

STIC-Biotech/ChemLib

70812

From: Sullivan, Daniel  
Sent: Sunday, July 14, 2002 10:52 AM  
To: STIC-Biotech/ChemLib  
Subject: CDB Search Request (10/053641)

For case number **10/053641**, please do a standard sequence search for SEQ ID NO: 1-8 against the commercial or interference nucleic acid databases or both.

Thanks very much.

Daniel M. Sullivan  
Examiner AU 1636  
Room: 12D12  
Mail Box: 11E12  
Tel: 703-305-4448

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 7/15/02  
Date Completed: 7/16/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: 8  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 01  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 17:12:34 ; Search time 2368.24 Seconds  
(without alignments)  
918.978 Million cell updates/sec

Title: US-10-053-641-1

Perfect score: 104

Sequence: 1 gatcctttatgtgtgtttac.....gcacaaatgcatcctgggc 104

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl :

1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pi.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----|-------------|
| -----      |       |       |        |    |    |             |

| 1    | 100  | 96.2 | 226  | 12 | SYNHIR     | M26762 Synthetic h |
|------|------|------|------|----|------------|--------------------|
| 2    | 96   | 92.3 | 129  | 6  | I41183     | I41183 Sequence 3  |
| 3    | 96   | 92.3 | 235  | 6  | I41187     | I41187 Sequence 7  |
| c 4  | 96   | 92.3 | 235  | 6  | I41188     | I41188 Sequence 8  |
| 5    | 96   | 92.3 | 726  | 6  | I41189     | I41189 Sequence 9  |
| c 6  | 96   | 92.3 | 726  | 6  | I41190     | I41190 Sequence 10 |
| 7    | 88.4 | 85.0 | 220  | 12 | SYNHIRMED  | M14964 Synthetic h |
| 8    | 88.4 | 85.0 | 224  | 6  | A00630     | A00630 Artificial  |
| c 9  | 88.4 | 85.0 | 224  | 6  | A00631     | A00631 Artificial  |
| 10   | 88.4 | 85.0 | 224  | 6  | A04277     | A04277 Artificial  |
| c 11 | 88.4 | 85.0 | 224  | 6  | A04278     | A04278 Artificial  |
| 12   | 88.4 | 85.0 | 224  | 6  | A07345     | A07345 Synthetic D |
| c 13 | 88.4 | 85.0 | 224  | 6  | A07346     | A07346 Synthetic D |
| 14   | 86.8 | 83.5 | 109  | 6  | A34234     | A34234 Synthetic h |
| c 15 | 86.8 | 83.5 | 109  | 6  | A34235     | A34235 Synthetic h |
| 16   | 86.8 | 83.5 | 217  | 6  | A03695     | A03695 Nucleotide  |
| c 17 | 86.8 | 83.5 | 217  | 6  | A03696     | A03696 Nucleotide  |
| 18   | 86.8 | 83.5 | 217  | 6  | A34238     | A34238 Synthetic d |
| c 19 | 86.8 | 83.5 | 217  | 6  | A34239     | A34239 Synthetic d |
| 20   | 86.8 | 83.5 | 217  | 6  | E00657     | E00657 DNA encodin |
| c 21 | 83.2 | 80.0 | 178  | 6  | A13978     | A13978 Synthetic D |
| 22   | 83.2 | 80.0 | 180  | 6  | A02448     | A02448 Nucleotide  |
| c 23 | 83.2 | 80.0 | 180  | 6  | A02449     | A02449 Nucleotide  |
| 24   | 83.2 | 80.0 | 180  | 6  | A13977     | A13977 Synthetic D |
| c 25 | 82.4 | 79.2 | 195  | 6  | A03693     | A03693 Nucleotide  |
| 26   | 82.4 | 79.2 | 195  | 6  | A03694     | A03694 Nucleotide  |
| c 27 | 82.4 | 79.2 | 279  | 6  | A01139     | A01139 Fuson DNA   |
| 28   | 82.4 | 79.2 | 279  | 6  | I26634     | I26634 Sequence 3  |
| c 29 | 82.2 | 79.0 | 238  | 6  | A04616     | A04616 HVI gene. 8 |
| 30   | 81   | 77.9 | 106  | 6  | I41186     | I41186 Sequence 6  |
| c 31 | 80.8 | 77.7 | 225  | 6  | A04613     | A04613 Nucleotide  |
| 32   | 80.8 | 77.7 | 225  | 6  | A04614     | A04614 Nucleotide  |
| c 33 | 80.8 | 77.7 | 226  | 6  | A04615     | A04615 Nucleotide  |
| 34   | 79   | 76.0 | 93   | 6  | I63540     | I63540 Sequence 3  |
| c 35 | 78.2 | 75.2 | 8491 | 6  | AR031529   | AR031529 Sequence  |
| 36   | 78   | 75.0 | 212  | 6  | A13382     | A13382 DNA sequenc |
| c 37 | 78   | 75.0 | 212  | 6  | A13383     | A13383 DNA sequenc |
| 38   | 78   | 75.0 | 212  | 6  | A34619     | A34619 Synthetic h |
| c 39 | 78   | 75.0 | 212  | 6  | A34620     | A34620 Synthetic h |
| 40   | 78   | 75.0 | 212  | 6  | E00711     | E00711 Synthetic D |
| c 41 | 77.4 | 74.4 | 4477 | 12 | SYNEXVECHD | M88535 Expression  |
| 42   | 74.2 | 71.3 | 223  | 6  | A19994     | A19994 SEQ ID NO:  |
| c 43 | 74.2 | 71.3 | 223  | 6  | A20035     | A20035 SEQ ID NO:  |
| 44   | 72.6 | 69.8 | 211  | 6  | E03003     | E03003 DNA encodin |
| 45   | 72   | 69.2 | 1467 | 6  | A20030     | A20030 SEQ ID NO:  |

#### ALIGNMENTS

| RESULT     | 1 | SYNHIR | SYNHIR  | 226 bp | DNA | linear | SYN 27-APR-1993 |
|------------|---|--------|---|--------|-----|--------|-----------------|
| LOCUS      |   |        | Synthetic hirudin gene, complete cds.   |        |     |        |                 |
| DEFINITION |   |        | Synthetic hirudin gene, complete cds.   |        |     |        |                 |
| ACCESSION  |   |        | M26762  |        |     |        |                 |
| VERSION    |   |        | M26762.1  |        |     |        |                 |
| KEYWORDS   |   |        | hirudin; proteinase inhibitor; thrombin inhibitor.  |        |     |        |                 |
| SOURCE     |   |        | Synthetic DNA.  |        |     |        |                 |
| ORGANISM   |   |        | artificial construct  |        |     |        |                 |
| REFERENCE  |   |        | 1 (bases 1 to 226)  |        |     |        |                 |
| AUTHORS    |   |        | Bergmann,C., Dodd,J., Koehler,S., Fink,E. and Gassen,H.G.   |        |     |        |                 |
| TITLE      |   |        | Chemical synthesis and expression of a gene coding for hirudin, the thrombin-specific inhibitor from the leech Hirudo medicinalis |        |     |        |                 |
| JOURNAL    |   |        | Biol. Chem. Hoppe-Seyler 367, 731-740 (1986)  |        |     |        |                 |
| MEDLINE    |   |        | 87026239  |        |     |        |                 |
| FEATURES   |   |        | Location/Qualifiers   |        |     |        |                 |
| source     |   |        | 1..226  |        |     |        |                 |
|            |   |        | /organism="synthetic construct"   |        |     |        |                 |
|            |   |        | /db_xref="taxon:32630"  |        |     |        |                 |
|            |   |        | 22..222   |        |     |        |                 |
|            |   |        | /note="hirudin"   |        |     |        |                 |
|            |   |        | /codon_start=1  |        |     |        |                 |
| CDS        |   |        |   |        |     |        |                 |

/transl\_table=11  
/protein\_id="AA72772.1"  
/db\_xref="GI:208479"  
/translation="MVVYDCTESGQNLCLCEGSNVCGGNKCILGSDGKFNOCVTGE  
GTPRPQSHNDGDFEIPPEYLQ"  
63 a 56 c 61 g 46 t  
BASE COUNT  
ORIGIN

Query Match 96.2%; Score 100; DB 12; Length 226;  
Best Local Similarity 100.0%; Pred. No. 1e-22;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctttatgtttttacactgactgactgaatccggtcagaacctgtgcctgtgcgaagg 64  
|||||  
Db 18 CTTTATGTTTACACTGACTGACTGAATCCGGTCAGAACCTGTGCTGTGCGAAG 77  
|||||

QY 65 ctctaacttttgcggcagggaacaatgcatcctgggc 104  
|||||

Db 78 CTCFAACGTTTGGCGCCAGGCAACAATGATCCTGGGC 117  
|||||

RESULT 2  
LOCUS I41183 129 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 3 from patent US 5624822.  
ACCESSION I41183  
VERSION I41183.1 GI:2081773  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 129)  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 3 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..129

BASE COUNT 36 a 35 c 29 g 29 t  
ORIGIN

Query Match 92.3%; Score 96; DB 6; Length 129;  
Best Local Similarity 100.0%; Pred. No. 2.1e-21;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atggtttttacactgactgactgaatccggtcagaacctgtgcctgtgcgaaggctct 68  
|||||  
Db 31 ATGGTTGTTTACACTGACTGACTGAATCCGGTCAGAACCTGTGCTGTGCGAAGGCTCT 90  
|||||

QY 69 aacgtttgcggccagggaacaatgcatcctgggc 104  
|||||  
Db 91 AACGTTTGGCGCCAGGCAACAATGATGATCCTGGGC 126  
|||||

RESULT 3  
LOCUS I41187 235 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 7 from patent US 5624822.  
ACCESSION I41187  
VERSION I41187.1 GI:2081777  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 235)  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 7 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..235

/organism="unknown"

BASE COUNT  
ORIGIN

Query Match 92.3%; Score 96; DB 6; Length 235;  
Best Local Similarity 100.0%; Pred. No. 2.2e-21;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atggtttttacactgactgactgaatccggtcagaacctgtgcctgtgcgaaggctct 68  
|||||  
Db 31 ATGGTTGTTTACACTGACTGACTGAATCCGGTCAGAACCTGTGCTGTGCGAAGGCTCT 90  
|||||

QY 69 aacgtttgcggccagggaacaatgcatcctgggc 104  
|||||  
Db 91 AACGTTTGGCGCCAGGCAACAATGATGATCCTGGGC 126  
|||||

RESULT 4  
LOCUS I41188 235 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 8 from patent US 5624822.  
ACCESSION I41188  
VERSION I41188.1 GI:2081778  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 235)  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 8 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..235

BASE COUNT 42 a 58 c 64 g 71 t  
ORIGIN

Query Match 92.3%; Score 96; DB 6; Length 235;  
Best Local Similarity 100.0%; Pred. No. 2.2e-21;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atggtttttacactgactgactgaatccggtcagaacctgtgcctgtgcgaaggctct 68  
|||||  
Db 209 ATGGTTGTTTACACTGACTGACTGAATCCGGTCAGAACCTGTGCTGTGCGAAGGCTCT 150  
|||||

QY 69 aacgtttgcggccagggaacaatgcatcctgggc 104  
|||||  
Db 149 AACGTTTGGCGCCAGGCAACAATGATGATCCTGGGC 114  
|||||

RESULT 5  
LOCUS I41189 726 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 9 from patent US 5624822.  
ACCESSION I41189  
VERSION I41189.1 GI:2081779  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 726)  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 9 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..726

BASE COUNT 281 a 171 c 126 g 148 t  
ORIGIN

Query Match 92.3%; Score 96; DB 6; Length 726;





```

KEYWORDS
SOURCE      synthetic construct.
ORGANISM    synthetic construct
REFERENCE    1 (bases 1 to 224)
AUTHORS
TITLE       DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR
            PREPARING SUCH PROTEIN
JOURNAL     Patent: WO 8603517-A 16 19-JUN-1986;
FEATURES    Location/Qualifiers
            source
            1..224
            /organism="synthetic construct"
            /db_xref="taxon:32630"
BASE COUNT  51 a 60 c 51 g 62 t
ORIGIN

Query Match      85.0%; Score 88.4; DB 6; Length 224;
Best Local Similarity 93.9%; Pred. No. 7.4e-19;
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 6 tttatggtgtttacactgactgcaatccgggtcagaacctgtgcctgtgcgaagc 65
Db 221 TCTATGTTGTTTACACTGACTGCACCGAATCTGGTCAGAACCTGTGCCTGTGCGAAGT 162

Qy 66 tctaacgtttgcgcaggcgaacaaatgcctcctggg 103
Db 161 TCTAACGTTTGGGTGTCAGGGAACAAATGCATCCTGGG 124

RESULT 10
LOCUS      A04277          224 bp DNA linear PAT 15-APR-1993
DEFINITION Artificial sequence for hirudin-like fusion protein.
ACCESSION  A04277
VERSION    A04277.1 GI:344867
KEYWORDS
SOURCE     synthetic construct.
ORGANISM   synthetic construct
REFERENCE   1 (bases 1 to 224)
AUTHORS
TITLE       DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR
            PREPARING SUCH PROTEIN
JOURNAL     Patent: WO 8603517-A 19 19-JUN-1986;
FEATURES    Location/Qualifiers
            source
            1..224
            /organism="synthetic construct"
            /db_xref="taxon:32630"
BASE COUNT  62 a 51 c 60 g 51 t
ORIGIN

Query Match      85.0%; Score 88.4; DB 6; Length 224;
Best Local Similarity 93.9%; Pred. No. 7.4e-19;
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 6 tttatggtgtttacactgactgcaatccgggtcagaacctgtgcctgtgcgaagc 65
Db 4 TCTATGTTGTTTACACTGACTGCACCGAATCTGGTCAGAACCTGTGCCTGTGCGAAGT 63

Qy 66 tctaacgtttgcgcaggcgaacaaatgcctcctggg 103
Db 64 TCTAACGTTTGGGTGTCAGGGAACAAATGCATCCTGGG 101

RESULT 11
LOCUS      A04278/c       224 bp DNA linear PAT 15-APR-1993
DEFINITION Artificial sequence for hirudin-like fusion protein, reverse
            complement.
ACCESSION  A04278
VERSION    A04278.1 GI:344868

```

```

KEYWORDS
SOURCE      synthetic construct.
ORGANISM    synthetic construct
REFERENCE    1 (bases 1 to 224)
AUTHORS
TITLE       DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR
            PREPARING SUCH PROTEIN
JOURNAL     Patent: WO 8603517-A 20 19-JUN-1986;
FEATURES    Location/Qualifiers
            source
            1..224
            /organism="synthetic construct"
            /db_xref="taxon:32630"
BASE COUNT  51 a 60 c 51 g 62 t
ORIGIN

Query Match      85.0%; Score 88.4; DB 6; Length 224;
Best Local Similarity 93.9%; Pred. No. 7.4e-19;
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 6 tttatggtgtttacactgactgcaatccgggtcagaacctgtgcctgtgcgaagc 65
Db 221 TCTATGTTGTTTACACTGACTGCACCGAATCTGGTCAGAACCTGTGCCTGTGCGAAGT 162

Qy 66 tctaacgtttgcgcaggcgaacaaatgcctcctggg 103
Db 161 TCTAACGTTTGGGTGTCAGGGAACAAATGCATCCTGGG 124

RESULT 12
LOCUS      A07345          224 bp DNA linear PAT 03-AUG-1993
DEFINITION Synthetic DNA for hirudin-similar protein.
ACCESSION  A07345
VERSION    A07345.1 GI:413049
KEYWORDS
SOURCE     synthetic construct.
ORGANISM   synthetic construct
REFERENCE   1 (bases 1 to 224)
AUTHORS
TITLE       DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR
            PREPARING SUCH PROTEIN
JOURNAL     Patent: DE 3445517-A 1 19-JUN-1986;
FEATURES    Location/Qualifiers
            source
            1..224
            /organism="synthetic construct"
            /db_xref="taxon:32630"
BASE COUNT  62 a 51 c 60 g 51 t
ORIGIN

Query Match      85.0%; Score 88.4; DB 6; Length 224;
Best Local Similarity 93.9%; Pred. No. 7.4e-19;
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 6 tttatggtgtttacactgactgcaatccgggtcagaacctgtgcctgtgcgaagc 65
Db 4 TCTATGTTGTTTACACTGACTGCACCGAATCTGGTCAGAACCTGTGCCTGTGCGAAGT 63

Qy 66 tctaacgtttgcgcaggcgaacaaatgcctcctggg 103
Db 64 TCTAACGTTTGGGTGTCAGGGAACAAATGCATCCTGGG 101

RESULT 13
LOCUS      A07346/c       224 bp DNA linear PAT 03-AUG-1993
DEFINITION Synthetic DNA for hirudin-similar protein
            (reverse complement).
ACCESSION  A07346
VERSION    A07346.1 GI:411427
KEYWORDS
SOURCE     synthetic construct.
ORGANISM   synthetic construct

```

artificial sequence.  
REFERENCE 1 (bases 1 to 224)  
AUTHORS  
JOURNAL  
FEATURES source  
Location/Qualifiers  
1. .224  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 51 a 60 c 51 g 62 t  
ORIGIN

Query Match 85.0%; Score 88.4; DB 6; Length 224;  
Best Local Similarity 93.9%; Pred. No. 7.4e-19;  
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 6 ttatggtgtttacactgactgactgaatccggcagaacctgtgcctgtgcgaagc 65  
DB 221 TCTATGTTGTTTACACTGACTGACCGAATCTGGTCAGAACCTGTGCCTGTGCGAAGT 162  
QY 66 tctaactgttgcggcaggcaggaacaaatgcatcctggg 103  
DB 161 TCTAACGTTTGGCGTCAGGGAACAAATGCATCCTGGG 124

RESULT 14  
A34234  
LOCUS A34234 109 bp DNA linear PAT 25-JUL-1996  
DEFINITION Synthetic hirudin analogue duplex I.  
ACCESSION A34234  
VERSION A34234.1 GI:1568164  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
artificial sequence.  
REFERENCE 1 (bases 1 to 109)  
AUTHORS Liersch, M., Rink, H., Maerki, W., Gruetter, M.G. and Meyhack, B.  
TITLE Process for the preparation of thrombin inhibitors  
JOURNAL Patent: EP 0168342-A 6 15-JAN-1986;  
CIBA-GEIGY AG; CIBA-GEIGY AG; UCP GEN-PHARMA AG  
FEATURES source  
1. .109  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 23 a 25 c 30 g 31 t  
ORIGIN

Query Match 83.5%; Score 86.8; DB 6; Length 109;  
Best Local Similarity 92.9%; Pred. No. 2.4e-18;  
Matches 91; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 6 ttatggtgtttacactgactgactgaatccggcagaacctgtgcctgtgcgaagc 65  
DB 7 TTCATGTTGTTTACACCGACTGACCGAATCTGGTCAGAACCTGTGCCTGTGCGAAGT 66  
QY 66 tctaactgttgcggcaggcaggaacaaatgcatcctggg 103  
DB 67 TCTAACGTTTGGCGTCAGGGAACAAATGCATCCTGGG 104

RESULT 15  
A34235/C  
LOCUS A34235 109 bp DNA linear PAT 25-JUL-1996  
DEFINITION Synthetic hirudin analogue duplex I.  
ACCESSION A34235  
VERSION A34235.1 GI:1568165  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
artificial sequence.  
REFERENCE 1 (bases 1 to 109)  
AUTHORS Liersch, M., Rink, H., Maerki, W., Gruetter, M.G. and Meyhack, B.

TITLE Process for the preparation of thrombin inhibitors  
JOURNAL Patent: EP 0168342-A 7 15-JAN-1986;  
CIBA-GEIGY AG; CIBA-GEIGY AG; UCP GEN-PHARMA AG  
FEATURES source  
Location/Qualifiers  
1. .109  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 31 a 30 c 25 g 23 t  
ORIGIN

Query Match 83.5%; Score 86.8; DB 6; Length 109;  
Best Local Similarity 92.9%; Pred. No. 2.4e-18;  
Matches 91; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 6 ttatggtgtttacactgactgactgaatccggcagaacctgtgcctgtgcgaagc 65  
DB 103 TTCATGTTGTTTACACCGACTGACCGAATCTGGTCAGAACCTGTGCCTGTGCGAAGT 44  
QY 66 tctaactgttgcggcaggcaggaacaaatgcatcctggg 103  
DB 43 TCTAACGTTTGGCGTCAGGGAACAAATGCATCCTGGG 6

Search completed: July 15, 2002, 22:28:32  
Job time: 18958 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 19:57:07 ; Search time 758.37 Seconds  
(without alignments)  
235.451 Million cell updates/sec

Title: US-10-053-641-1  
 Perfect score: 104  
 Sequence: 1 gatccttatggttggtttac.....gcaacaatgcacctgggc 104

Scoring table: IDENTITY\_NUC

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802:\*

|     |  |
|-----|--|
| 1:  | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*  |
| 2:  | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*  |
| 3:  | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*  |
| 4:  | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*  |
| 5:  | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*  |
| 6:  | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*  |
| 7:  | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*  |
| 8:  | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*  |
| 9:  | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*  |
| 10: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*  |
| 11: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*  |
| 12: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*  |
| 13: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*  |
| 14: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*  |
| 15: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*  |
| 16: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*  |
| 17: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*  |
| 18: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*  |
| 19: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*  |
| 20: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*  |
| 21: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*  |
| 22: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.* |
| 23: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.* |
| 24: | /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*  |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 96    | 92.3        | 726    | 12    | AAQ12379    |
| 2          | 93    | 89.4        | 468    | 16    | AAQ93244    |
| 3          | 88.4  | 85.0        | 214    | 7     | AAQ60746    |
| 4          | 86.8  | 83.5        | 217    | 7     | AAQ60355    |
| 5          | 86.8  | 83.5        | 217    | 8     | AAQ70323    |
| 6          | 82.4  | 79.2        | 196    | 8     | AAQ70319    |
| 7          | 82.4  | 79.2        | 197    | 12    | AAQ13831    |
| 8          | 80.8  | 77.7        | 229    | 9     | AAQ81294    |
| 9          | 79    | 76.0        | 238    | 9     | AAQ81295    |

|    |      |      |      |    |          |
|----|------|------|------|----|----------|
| 10 | 78.2 | 75.2 | 227  | 22 | AAE61507 |
| 11 | 72.6 | 69.8 | 210  | 12 | AAQ10452 |
| 12 | 72   | 69.2 | 1467 | 12 | AAQ12490 |
| 13 | 71.2 | 68.5 | 201  | 12 | AAQ12153 |
| 14 | 71.2 | 68.5 | 201  | 12 | AAQ25184 |
| 15 | 71.2 | 68.5 | 420  | 12 | AAQ12155 |
| 16 | 71.2 | 68.5 | 1458 | 12 | AAQ12162 |
| 17 | 69.6 | 66.9 | 198  | 13 | AAQ24986 |
| 18 | 69.6 | 66.9 | 198  | 13 | AAQ24986 |
| 19 | 68.8 | 66.2 | 231  | 7  | AAN60354 |
| 20 | 66.4 | 63.8 | 304  | 15 | AAQ54996 |
| 21 | 65.8 | 63.3 | 567  | 15 | AAQ54997 |
| 22 | 64.8 | 62.3 | 195  | 16 | AAQ81528 |
| 23 | 64.8 | 62.3 | 257  | 10 | AAN91836 |
| 24 | 64.8 | 62.3 | 257  | 10 | AAN91867 |
| 25 | 64.8 | 62.3 | 1082 | 15 | AAQ64146 |
| 26 | 64.8 | 62.3 | 1082 | 15 | AAQ45312 |
| 27 | 64.8 | 62.3 | 1130 | 16 | AAQ81527 |
| 28 | 64.6 | 62.1 | 219  | 15 | AAQ54995 |
| 29 | 63   | 60.6 | 207  | 12 | AAQ10453 |
| 30 | 63   | 60.6 | 2115 | 19 | AAV44279 |
| 31 | 62.6 | 60.2 | 198  | 15 | AAQ63876 |
| 32 | 59.4 | 57.1 | 1096 | 20 | AAV83966 |
| 33 | 58.6 | 56.3 | 183  | 12 | AAQ14927 |
| 34 | 58.6 | 56.3 | 185  | 12 | AAQ14926 |
| 35 | 56.4 | 54.2 | 195  | 13 | AAQ27889 |
| 36 | 56.4 | 54.2 | 198  | 16 | AAQ82762 |
| 37 | 56.4 | 54.2 | 204  | 16 | AAQ82764 |
| 38 | 54.2 | 52.1 | 379  | 6  | AAN50395 |
| 39 | 54.2 | 52.1 | 379  | 7  | AAN60043 |
| 40 | 54.2 | 52.1 | 381  | 6  | AAN50396 |
| 41 | 52.6 | 50.6 | 204  | 14 | AAQ38407 |
| 42 | 52.6 | 50.6 | 204  | 16 | AAQ83521 |
| 43 | 52   | 50.0 | 526  | 11 | AAQ03237 |
| 44 | 49   | 47.1 | 195  | 13 | AAQ27601 |
| 45 | 49   | 47.1 | 276  | 13 | AAQ27600 |

## ALIGNMENTS

|          |  |                                 |
|----------|--|---------------------------------|
| RESULT   | 1                                      |                                 |
| AAQ12379 |  |                                 |
| ID       | AAQ12379                               | standard; DNA; 726 BP.          |
| XX       |  |                                 |
| AC       | AAQ12379;                              |                                 |
| XX       |  |                                 |
| DT       | 17-SEP-1991                            | (first entry)                   |
| XX       |  |                                 |
| DE       | Hirudin peptide/Protein A fusion gene. |                                 |
| XX       |  |                                 |
| XX       | anticoagulant; fusion protein; ds.     |                                 |
| KW       |  |                                 |
| XX       |  |                                 |
| FH       | Key                                    | Location/Qualifiers             |
| FT       | mat_peptide                            | 1..489                          |
| FT       |  | /*tag= a                        |
| FT       |  | /product= Protein A             |
| FT       | mat_peptide                            | 529..720                        |
| FT       |  | /*tag= b                        |
| FT       |  | /product= Hirudin               |
| FT       | misc_RNA                               | 490..528                        |
| FT       |  | /*tag= c                        |
| FT       |  | /product= linker oligopeptide y |
| XX       |  |                                 |
| PN       | DE3942580-A.                           |                                 |
| XX       |  |                                 |
| PD       | 27-JUN-1991.                           |                                 |
| XX       |  |                                 |
| PF       | 22-DEC-1989;                           | 89DE-3942580.                   |
| XX       |  |                                 |
| PR       | 22-DEC-1989;                           | 89DE-3942580.                   |
| XX       |  |                                 |
| PA       | (BADI ) BASF AG.                       |                                 |

```

XX PI Korwer W;
XX WPI; 1991-194236/27.
XX P-PSDB; AAR12751.
XX Hirudin peptide prodn. by cleaving new fusion peptide - of
XX hirudin and protein A, expressed in high yield and stable,
XX soluble form by transformed E.coli
XX Example 1; Page 6-7; 9pp; German.
XX This sequence is an example of a fusion construct for expression of
XX the fusion peptide of the invention. The Met residue in the linker
XX oligopeptide allows cleavage by CNBr to release two fragments which
XX can be easily separated by IFC affinity chromatography. (The protein
XX A component binds to IFC sepharose). Increased yields of hirudin are
XX obtained using this fusion construct.
XX Sequence 726 BP; 280 A; 171 C; 126 G; 149 T; 0 other;

Query Match          92.3%; Score 96; DB 12; Length 726;
Best Local Similarity 100.0%; Pred. No. 3.2e-24;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atggtgtttacactgactgactgaatccggtcagaaacctgtgcctgtgcgaagctct 68
DB 523 atggtgtttacactgactgactgaatccggtcagaaacctgtgcctgtgcgaagctct 582
QY 69 aacgtttgcccaggcggaacaaatgcattcctgggc 104
DB 583 aacgtttgcccaggcggaacaaatgcattcctgggc 618

RESULT 2
AAQ93244
ID AAQ93244 standard; DNA; 468 BP.
AC AAQ93244;
XX
XX 01-NOV-1995 (first entry)
XX Fusion construct of glucoamylase-hirudin DNA.
XX expression cassette; recombinant protein; production; Hansenula;
XX yeast; leader sequence; adaptor; alpha helix; glucoamylase;
XX secretion; processing; thrombin inhibitor; hirudin; ds.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX primer_bind complement (1..33)
XX CDS 13..444
XX /*tag= a
XX /*tag= b
XX /*tag= c
XX primer_bind /product= glucoamylase-hirudin fusion protein
XX 199..223
XX /*tag= c
XX primer_bind /note= "primer AAQ85841 binding site"
XX complement (220..267)
XX /*tag= d
XX /*tag= "primer AAQ85842 binding site"
XX 13..228
XX /*tag= e
XX /*tag= "encodes amino acids 1-72 of glucoamylase"
XX misc_feature 241..246
XX /*tag= f
XX /*tag= "encodes processor signal"
XX 247..444
XX /*tag= g
XX /*tag= "encodes hirudin-h120"

```

```

FT primer_bind 0..468
FT /*tag= h
FT /note= "primer AAQ85843 binding site"
XX
XX DE4329969-A.
XX
XX 09-MAR-1995.
XX
XX 04-SEP-1993; 93DE-4329969.
XX
XX 04-SEP-1993; 93DE-4329969.
XX
XX (BADI ) BASF AG.
XX (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
XX
XX Bollschweiler C, Janowicz ZA, Piontek M, Schweden J;
XX Strasserrawn, Weydemann U;
XX
XX WPI; 1995-107810/15.
XX P-PSDB; AAR76951.
XX
XX Recombinant protein prodn. in Hansenula yeast - transformed with
XX expression cassette contg. leader, adaptor, processing signal and
XX gene, provides efficient secretion and correct processing
XX Example 1; Fig 1; 10pp; German.
XX
XX AAQ93244 is a DNA fusion product resulting from ligation of an
XX EcoRI-PvuI fragment (see AAQ85840-41) and a PvuI-SalI fragment, encoding
XX a leader sequence, an adaptor (see AAR71472, amino acids 23-72 of GAM
XX (glucamylase from Schwanniomycetes occidentalis, plus -His-Pro-Leu-Gln at
XX the C-terminal end), a processor peptide (Lys-Arg) and a structural gene,
XX encoding hirudin. AAQ93244 is an example of an expression vector insert
XX of the general formula: L-A-P-Gene (L = leader sequence; A = adaptor;
XX P = processor; Gene = structural gene). The cassettes ensure efficient
XX secretion and correct processing of heterologous structural genes in
XX yeast of the genus Hansenula, and so provides high yields of mature
XX proteins and facilitates subsequent purification.
XX
XX Sequence 468 BP; 126 A; 110 C; 112 G; 120 T; 0 other;

Query Match          89.4%; Score 93; DB 16; Length 468;
Best Local Similarity 100.0%; Pred. No. 3.3e-23;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gttgtttacactgactgactgaatccggtcagaaacctgtgcctgtgcgaagctctaac 71
DB 247 gttgtttacactgactgactgaatccggtcagaaacctgtgcctgtgcgaagctctaac 306
QY 72 gttgtgtccaggcggaacaaatgcattcctgggc 104
DB 307 gttgtgtccaggcggaacaaatgcattcctgggc 339

RESULT 3
AAQ60746
ID AAQ60746 standard; DNA; 224 BP.
XX
XX AAQ60746;
XX
XX 01-JAN-1980 (first entry)
XX
XX DNA encoding hirudin-like polypeptide.
XX
XX Hirudin; thrombin-antagonist; anticoagulant; ds.
XX Synthetic.
XX
XX DE3445517-A.
XX
XX 19-JUN-1986.

```







|  |  |
|--|--|
| KW   | Anticoagulant; thrombin inhibitor; yeast expression cassette; ss.        |
| XX   | Synthetic.   |
| OS   |  |
| XX   |  |
| FFH  | Key Location/Qualifiers  |
| FT   | CDS 1..237   |
| FT   | /tag= a  |
| XX   |  |
| XX   | EP252854-A.  |
| PN   |  |
| PD   | 13-JAN-1988.   |
| XX   |  |
| PPF  | 10-JUL-1987; 87EP-0401649.   |
| XX   |  |
| PR   | 01-DEC-1986; 86FR-0016722.   |
| PR   | 01-JUL-1986; 86FR-0010090.   |
| XX   |  |
| PA   | (TRAN-) TRANSGENE SA.  |
| XX   |  |
| PI   | Labat N, Loison G, Balland A, Lemoine Y;                                 |
| XX   |  |
| DR   | WPI; 1988-008792/02.   |
| DR   | P-PADB; AAP80995.  |
| XX   |  |
| PPT  | New DNA block providing expression of hirudin in yeast -                 |
| PPT  | comprises hirudin gene, signal and leader sequences and sequence         |
| PPT  | for unique proteolytic cleavage site                                     |
| XX   |  |
| PS   | Disclosure; ; pp: French.  |
| XX   |  |
| CCC  | The patent is for a new DNA block providing expression of hirudin (H) in |
| CCC  | yeast. Also new are plasmids contg. the block plus at least one yeast    |
| CCC  | origin of replication and yeasts, esp. Saccharomyces cerevisiae,         |
| CCC  | transformed with such plasmids.  |
| XX   |  |
| SQ   | Sequence 238 BP; 79 A; 47 C; 55 G; 57 T; 0 other;                        |
|  |  |
| Query Match 76.0%; Score 79; DB 9; Length 238;               |  |
| Best Local Similarity 89.5%; Pred.No. 2.4e-18;               |  |
| Matches 85; Conservative 0; Mismatches 10; Indels 0; Gaps 0; |  |
| QY   | 9 atgggttttacactgcactgcactgaatccggtcacagaacctgtgcctgtcgaaagtctt 68<br>   |
| Ddb  | 1 atggtagtttataccgactgcaccgaatccggtcacagaacctgtgcctgtcgaaagtctt 60<br>   |
| QY   | 69 aacgttttgccaggcggcaaaaatgcattctggg 103<br>                            |
| Ddb  | 61 aacgtttgcggtcagggttaaaaaatgcattctgg 95<br>                            |
| RESULT 10  |  |
| AAAF61507  |  |
| ID   | AAAF61507 standard; DNA; 227 BP.   |
| XX   |  |
| AC   | AAAF61507;   |
| XX   |  |
| DT   | 18-JUN-2001 (first entry)  |
| XX   |  |
| DE   | S. marcescens hirudin DNA fragment.                                      |
| XX   |  |
| KW   | Hirudin; outer membrane protein; oprF; lambdaB; fumarate reductase;      |
| KW   | Leu-hirudin; Leu-Thr2-63-desulfato-hirudin; antithrombotic; ds.          |
| XX   |  |
| OS   | Serratia marcescens.   |
| XX   |  |
| PN   | DE19944870-A1.   |
| XX   |  |
| PD   | 29-MAR-2001.   |
| XX   |  |
| PF   | 18-SEP-1999; 99DF-1044870.   |
| XX   |  |
| PR   | 18-SEP-1999; 99DF-1044870.   |



PF 07-DEC-1990; 90WO-GB01911.  
 XX 07-DEC-1990; 90WO-GB01911.  
 PR 07-DEC-1989; 89GB-0027722.  
 XX (BRBI-) BRIT BIO-TECHN LTD.  
 PA Dawson KM, Hunter MG, Czaplewski LG;  
 XX WPI: 1991-208151/28.  
 PI P-PSDB; AAR12522.  
 XX Fusion protein cleavage by blood clotting enzyme - for prodn. of  
 XX fractions having greater antithrombotic activity for therapy and  
 XX prophylaxis.  
 XX Disclosure; Page 98; 115pp; English.  
 XX The sequence of the synthetic hirudin HV-1 gene was designed  
 CC based on the published amino acid sequence (Dodd J., et al FEBS  
 CC Letters 165 180 (1984)). The sequence of streptokinase was obtd.  
 CC from PCR amplified chromosomal DNA from S. equisimilis ATCC 10009  
 CC or ATCC 9642. The primers used for the PCR were based on the pub-  
 CC lished DNA sequence of S. equisimilis strain H46A (Malke, H., Roe,  
 CC B., and Ferretti, J.J., Gene 34 357-362 [1985]). The two  
 CC sequences were used to construct an expression vector in which the  
 CC streptokinase gene is linked to the hirudin gene via a linking  
 CC sequence encoding a cleavage site for factor Xa. The factor Xa is  
 CC present at the site of the target thrombus so the active agents are  
 CC released specifically at the place where clot formation is occurring.  
 CC See also AAQ12153-Q12156 and AAQ12158-Q12162.  
 XX Sequence 1467 BP; 494 A; 317 C; 292 G; 364 T; 0 other;  
 SQ

Query Match 69.2%; Score 72; DB 12; Length 1467;  
 Best Local Similarity 84.4%; Pred. No. 1.3e-15;  
 Matches 81; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 8 tatgtgtttacactgactgactgaatccggtcagaaacctgtgcctgtgcgaaggctc 67  
 Db 1251 tagagtgtttacacgactgactgaatccggtcagaaacctgtgtgtgtgaggttc 1310  
 QY 68 taacgtttggtccagggcacaataatgcattcctggg 103  
 Db 1311 taacgtctgtgtcagggtaacaaatgcattcctggg 1346

RESULT 13  
 AAQ12153  
 ID AAQ12153 standard; DNA; 201 BP.  
 XX AAQ12153;  
 AC AAQ12153;  
 XX 17-SEP-1991 (first entry)  
 DT Synthetic hirudin type HV-1 gene.  
 DE Fusion protein; blood clotting; coagulation; fibrinolysis;  
 KW antithrombotic; thrombolysis; streptokinase; ds.  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 FH 1..201  
 FT CDS /\*tag= a  
 FT  
 XX W09109125-A.  
 PN  
 XX 27-JUN-1991.  
 PD  
 XX 07-DEC-1990; 90WO-GB01911.  
 PF  
 XX

PR 07-DEC-1990; 90WO-GB01911.  
 PR 07-DEC-1989; 89GB-0027722.  
 XX (BRBI-) BRIT BIO-TECHN LTD.  
 XX Dawson KM, Hunter MG, Czaplewski LG;  
 XX WPI: 1991-208151/28.  
 DR P-PSDB; AAR12887.  
 XX Fusion protein cleavage by blood clotting enzyme - for prodn. of  
 XX fractions having greater antithrombotic activity for therapy and  
 XX prophylaxis  
 XX Disclosure; Page 68; 115pp; English.  
 XX The sequence was designed based on the published amino acid se-  
 CC quence (Dodd J., et al FEBS letters 165 180 (1984)). Unique re-  
 CC striction sites were incorporated to facilitate subsequent genetic  
 CC manipulation. The codons selected were those favourable for S.  
 CC cerevisiae or E. coli. The sequence was divided into 12 oligomers  
 CC which were synthesised and then annealed. The ligation prod. was  
 CC ligated to HindIII and EcoRI treated pUC19 plasmid DNA and the  
 CC resulting vector used to transform E. coli K12 HB87. plasmid pUC19  
 CC HV-1 was isolated from transformants and inserted into plasmid  
 CC pSW6, a shuttle vector, for expression. The gene can be used to  
 CC construct expression vectors in which the hirudin gene is linked to  
 CC a second gene encoding e.g. another hirudin protein, streptokinase  
 CC or a streptokinase-like protein, via a linking peptide. This pep-  
 CC tide link contains a cleavage site for e.g. factor X or thrombin  
 CC which can be cleaved, releasing the individual proteins which have  
 CC antithrombotic activity. The enzymes which cleave the fusion pro-  
 CC tein are present at the site of the target thrombus so the active  
 CC agents are released specifically at the place where clot formation  
 CC is occurring.  
 CC See also AAQ12154-Q12156, AAQ12158-Q12162 and AAQ12490.  
 XX Sequence 201 BP; 59 A; 43 C; 52 G; 47 T; 0 other;  
 SQ

Query Match 68.5%; Score 71.2; DB 12; Length 201;  
 Best Local Similarity 85.9%; Pred. No. 1.3e-15;  
 Matches 79; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 12 gttgtttacactgactgactgaatccggtcagaaacctgtgcctgtgcgaaggctc 71  
 Db 1 gttgtttacactgactgactgaatccggtcagaaacctgtgtgtgtgaggttc 60

QY 72 gttgtggtccagggcacaataatgcattcctggg 103  
 Db 61 gttgtggtcagggtaacaaatgcattcctggg 92

RESULT 14  
 AAQ25184  
 ID AAQ25184 standard; DNA; 201 BP.  
 XX AAQ25184;  
 AC AAQ25184;  
 XX 18-NOV-1992 (first entry)  
 DT Hirudin HV-1.  
 DE Restriction sites; Saccharomyces cerevisiae; Escherichia coli;  
 KW thrombin; fibrin; fibrinogen; ds.  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 FH 1..195  
 FT CDS /\*tag= a  
 FT  
 XX W09207874-A.  
 PN



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 17:23:54 ; Search time 174.99 Seconds  
(without alignments)  
145.985 Million cell updates/sec

Title: US-10-053-641-1

perfect score: 104  
Sequence: 1 qatcctttatggtttttac.....gcacaaatgcatacctgggc 104

Scoring table:      IDENTITY\_NUC  
Gapop 10.0      Gapext 1.0

Searched: 383533 seqs, 122816752 residues  
Total number of hits satisfying chosen parameters: 767066

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : Issued_Patents_NA.*
1: /cgcn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgcn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgcn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgcn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgcn2_6/ptodata/1/ina/PCtUS_COMB.seq.*
6: /cgcn2_6/ptodata/1/ina/backfiles1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID                | Description        |
|------------|-------|-------------|--------|----|-------------------|--------------------|
| 1          | 96    | 92.3        | 129    | 1  | US-08-262-384A-3  | Sequence 3, Appli  |
| 2          | 96    | 92.3        | 235    | 1  | US-08-262-384A-7  | Sequence 7, Appli  |
| C 3        | 96    | 92.3        | 235    | 1  | US-08-262-384A-8  | Sequence 8, Appli  |
| 4          | 96    | 92.3        | 726    | 1  | US-08-262-384A-9  | Sequence 9, Appli  |
| C 5        | 96    | 92.3        | 726    | 1  | US-08-262-384A-10 | Sequence 10, Appli |
| 6          | 86.8  | 83.5        | 109    | 6  | 5422249-6         | Patent No. 5422249 |
| 7          | 86.8  | 83.5        | 217    | 6  | 5422249-13        | Patent No. 5422249 |
| 8          | 86.8  | 83.5        | 218    | 6  | 5422249-4         | Patent No. 5422249 |
| 9          | 82.4  | 79.2        | 279    | 1  | US-08-186-222-3   | Sequence 3, Appli  |
| C 10       | 81    | 77.9        | 106    | 1  | US-08-262-384A-6  | Sequence 6, Appli  |
| 11         | 79    | 76.0        | 93     | 1  | US-07-859-453E-3  | Sequence 3, Appli  |
| C 12       | 78.2  | 75.2        | 8491   | 2  | US-08-757-439-1   | Sequence 1, Appli  |
| 13         | 78    | 75.0        | 212    | 6  | 5180668-10        | Patent No. 5180668 |
| 14         | 78    | 75.0        | 227    | 6  | US-07-983-064-8   | Sequence 8, Appli  |
| 15         | 77.4  | 74.4        | 227    | 6  | 5164304-3         | Patent No. 5164304 |
| 16         | 72    | 69.2        | 1467   | 1  | US-07-854-596B-46 | Sequence 46, Appli |
| 17         | 71.2  | 68.5        | 201    | 1  | US-07-854-596B-1  | Sequence 1, Appli  |
| 18         | 71.2  | 68.5        | 223    | 1  | US-07-854-596B-7  | Sequence 7, Appli  |
| 19         | 71.2  | 68.5        | 420    | 1  | US-07-854-596B-8  | Sequence 8, Appli  |
| C 20       | 71.2  | 68.5        | 1458   | 1  | US-07-854-596B-42 | Sequence 42, Appli |
| 21         | 69.6  | 66.9        | 198    | 1  | US-07-910-538-2   | Sequence 2, Appli  |
| C 22       | 69.6  | 66.9        | 198    | 1  | US-08-348-972-2   | Sequence 2, Appli  |
| 23         | 69.6  | 66.9        | 198    | 2  | US-08-861-459-2   | Sequence 2, Appli  |
| 24         | 66.4  | 63.8        | 195    | 1  | US-08-044-506B-28 | Sequence 28, Appli |
| C 25       | 66.4  | 63.8        | 206    | 1  | US-08-044-506B-21 | Sequence 21, Appli |
| 26         | 66.4  | 63.8        | 228    | 6  | 5179196-3         | Patent No. 5179196 |
| C 27       | 65.8  | 63.3        | 563    | 1  | US-08-044-506B-22 | Sequence 22, Appli |

|    |      |      |      |   |                   |                    |
|----|------|------|------|---|-------------------|--------------------|
| 28 | 64.8 | 62.3 | 195  | 1 | US-08-044-506B-27 | Sequence 27, Appl  |
| 29 | 64.8 | 62.3 | 195  | 1 | US-08-408-948A-7  | Sequence 7, Appl   |
| 30 | 64.8 | 62.3 | 202  | 1 | US-08-044-506B-20 | Sequence 20, Appl  |
| 31 | 64.8 | 62.3 | 1082 | 1 | US-08-715-252-1   | Sequence 1, Appl   |
| 32 | 64.8 | 62.3 | 1082 | 2 | US-08-453-051-3   | Sequence 3, Appl   |
| 33 | 64.8 | 62.3 | 1130 | 1 | US-08-408-948A-5  | Sequence 5, Appl   |
| 34 | 63   | 60.6 | 2115 | 2 | US-08-767-026-3   | Sequence 3, Appl   |
| 35 | 62.6 | 60.2 | 68   | 2 | US-08-861-459-3   | Sequence 3, Appl   |
| 36 | 58.4 | 56.2 | 198  | 2 | US-08-560-098A-38 | Sequence 38, Appl  |
| 37 | 56.4 | 54.2 | 195  | 1 | US-07-843-089E-46 | Sequence 46, Appl  |
| 38 | 56.4 | 54.2 | 195  | 1 | US-08-264-485-46  | Sequence 46, Appl  |
| 39 | 54.2 | 52.1 | 64   | 6 | 5422249-1         | Patent No. 5422249 |
| 40 | 52.6 | 50.6 | 204  | 2 | US-08-492-343-1   | Sequence 1, Appl   |
| 41 | 52.6 | 50.6 | 204  | 3 | US-09-080-865-2   | Sequence 2, Appl   |
| 42 | 52.6 | 50.6 | 204  | 5 | PCT-US94-10048-1  | Sequence 1, Appl   |
| 43 | 51.6 | 49.6 | 201  | 6 | 5422249-3         | Patent No. 5422249 |
| 44 | 49   | 47.1 | 195  | 6 | 5166318-14        | Patent No. 5166318 |
| 45 | 49   | 47.1 | 276  | 6 | 5166318-15        | Patent No. 5166318 |

## ALIGNMENTS

```

1
RESULT
US-08-262-384A-3
; Sequence 3, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: 530
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/Ep90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-262-384A-3

```

Query Match 92.3%; Score 96; DB 1; Length 129;  
Best Local Similarity 100.0%;  
Pred. No. 8.7e-26;

Qy 9 atggttggttacactgactgactgaatccggtcagaacctgtgcctgtgcgaaggctct 68  
|||||  
31 atgcttctttacactgactgactgaatccggtcagaacctgtgcctgtgcgaaggctct 90  
|||||



; APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 726 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-262-384A-9

Query Match 92.3%; Score 96; DB 1; Length 726;  
Best Local Similarity 100.0%; Pred. No. 1.5e-25;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atggttggttacactgactgactgaatccggtgcagaacctgtgctgtgcgaaggtct 68  
|||||  
DB 523 ATGTTGTTTACACTGACGTCACCTGAATCCGGTCAGAACCTGTGCTGTGCGAAGGCTCT 582  
|||||  
QY 69 aacgtttgcgcagggcgaacaaatgcattcctggc 104  
|||||  
DB 583 AACGTTTGGCCAGGCGCAACAAATGATCCTTGGGC 618  
|||||

RESULT 5  
US-08-262-384A-10/C  
; Sequence 10, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kell & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: 530  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 726 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-262-384A-10

Query Match 92.3%; Score 96; DB 1; Length 726;  
Best Local Similarity 100.0%; Pred. No. 1.5e-25;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atggttggttacactgactgactgaatccggtgcagaacctgtgctgtgcgaaggtct 68  
|||||  
DB 204 ATGTTGTTTACACTGACTGCTGAATCCGGTCAGAACCTGTGCTGTGCGAAGGCTCT 145  
|||||

QY 69 aacgtttgcgcagggcgaacaaatgcattcctggc 104  
|||||  
DB 144 AACGTTTGGCCAGGCGCAACAAATGATCCTTGGGC 109  
|||||

RESULT 6  
5422249-6  
; Patent No. 5422249  
; APPLICANT: LIERSCH, MANFRED; RINK, HANS; MARKI, WALTER; GRUTTER,  
; MARKUS G.; MEYHACK, BERND  
; TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF THROMBIN  
; INHIBITORS  
; NUMBER OF SEQUENCES: 15  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/121,974  
; FILING DATE: 15-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 700,9978  
; FILING DATE: 10-MAY-1991  
; APPLICATION NUMBER: 582,816  
; FILING DATE: 13-SEP-1990  
; APPLICATION NUMBER: 211,065  
; FILING DATE: 20-JUN-1988  
; APPLICATION NUMBER: 744,453  
; FILING DATE: 13-JUN-1985  
; SEQ ID NO: 6:  
; LENGTH: 109  
5422249-6

Query Match 83.5%; Score 86.8; DB 6; Length 109;  
Best Local Similarity 92.9%; Pred. No. 1.7e-22;  
Matches 91; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 ttatggtgtttacactgactgactgaatccggtgcagaacctgtgctgtgcgaaggc 65  
|||||  
DB 7 ttcactggtttacaccgactgcacgaatctggtcagaacctgtgctgtgcgaaggt 66  
|||||  
QY 66 tctaacgtttgcgcagggcgaacaaatgcattcctggg 103  
|||||  
DB 67 tctaacgtttgcgttcagggttaacaaatgcattcctggg 104  
|||||

RESULT 7  
5422249-13  
; Patent No. 5422249  
; APPLICANT: LIERSCH, MANFRED; RINK, HANS; MARKI, WALTER; GRUTTER,  
; MARKUS G.; MEYHACK, BERND  
; TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF THROMBIN  
; INHIBITORS  
; NUMBER OF SEQUENCES: 15  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/121,974  
; FILING DATE: 15-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 700,9978  
; FILING DATE: 10-MAY-1991  
; APPLICATION NUMBER: 582,816  
; FILING DATE: 13-SEP-1990  
; APPLICATION NUMBER: 211,065  
; FILING DATE: 20-JUN-1988  
; APPLICATION NUMBER: 744,453  
; FILING DATE: 13-JUN-1985  
; SEQ ID NO: 13:  
; LENGTH: 217  
5422249-13

Query Match 83.5%; Score 86.8; DB 6; Length 217;  
Best Local Similarity 92.9%; Pred. No. 2.1e-22;  
Matches 91; Conservative 0; Mismatches 7; Indels 0; Gaps 0;





;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/EP90/02084  
;; FILING DATE: 04-DEC-1990  
;; APPLICATION NUMBER: US 07861820  
;; FILING DATE: 18-JUN-1992  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 106 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-08-262-384A-6

Query Match 77.9%; Score 81; DB 1; Length 106;  
Best Local Similarity 100.0%; Pred. No. 2e-20;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 atggtttttacactgactgaatccggtcagaacctgtgctgtggaagctct 68  
|||  
Db 81 ATGGTTGTTTACACTGACTGCTGAATCCGGTCAGACCTGTGCTGTGCGAAGCTCT 22  
|||  
Qy 69 aacgcttgcggcagggaac 89  
|||  
Db 21 AACGTTTGGCGCAGGGCAAC 1  
|||

RESULT 11  
US-07-859-453E-3  
; Sequence 3, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:  
; APPLICANT: Kurfuerst, Manfred  
; APPLICANT: Ruebsamen, Klaus  
; APPLICANT: Schmied, Bernhard  
; APPLICANT: Koerwer, Wolfgang  
; APPLICANT: Schweden, Juergen  
; APPLICANT: Hoefken, Hans Wolfgang  
; TITLE OF INVENTION: Hitudin/polyalkylene glycol  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS version 7.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/859,453E  
FILING DATE: 29-MAY-1992  
CLASSIFICATION: 514  
CLASSIFICATION: A 61 K 37/64  
CLASSIFICATION: A 61 K 47/48  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP90/01998  
FILING DATE: 22-NOV-1990  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
US-07-859-453E-3

Query Match 76.0%; Score 79; DB 1; Length 93;  
Best Local Similarity 100.0%; Pred. No. 1e-19;

Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 8 tatgtttttacactgactgaatccggtcagaacctgtgctgtggaagctc 67  
|||  
Db 15 TATGGTTGTTTACACTGACTGCTGAATCCGGTCAGACCTGTGCTGTGCGAAGCTC 74  
|||  
Qy 68 taacgttttgcggcagggc 86  
|||  
Db 75 TAAGTTTGGCGCAGGGC 93  
|||

RESULT 12  
US-08-757-439-1/c  
; Sequence 1, Application US/08757439  
; Patent No. 5866371  
; GENERAL INFORMATION:  
; APPLICANT: BADZIONG, Werner  
; APPLICANT: HABERMANN, Paul  
; APPLICANT: MOELLER, Joerg  
; APPLICANT: ARETZ, Werner  
; TITLE OF INVENTION: PROCESS FOR USING THE YEAST ADH II  
; TITLE OF INVENTION: PROMOTER SYSTEM FOR THE PRODUCTION OF HETEROLOGOUS  
; TITLE OF INVENTION: PROMOTER SYSTEM FOR THE PRODUCTION OF HETEROLOGOUS  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/757,439  
FILING DATE: 27-NOV-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 19544233.4  
FILING DATE: 28-NOV-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: SANDERCOCK, Colin G.  
REGISTRATION NUMBER: 31,298  
REFERENCE/DOCKET NUMBER: 18748/303/HOCE  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8491 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-757-439-1

Query Match 75.2%; Score 78.2; DB 2; Length 8491;  
Best Local Similarity 91.2%; Pred. No. 8.8e-19;  
Matches 83; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
Qy 13 ttgtttacactgactgaatccggtcagaacctgtgctgtggaaggtctaacg 72  
|||  
Db 6986 TTACGTATACTGACTGCTGAATCTGCTCAGAACCTGTGCTGTGCGAAGGATCTAACG 6927  
|||  
Qy 73 ttgtgcggcaggccaacaaatgcatcctggg 103  
|||  
Db 6926 TTTGCGCCAGGGTAACAAATGATCCTTGG 6896  
|||

```

;          TOPOLOGY: linear
;          MOLECULE TYPE: other DNA
US-07-982-064-8

Query Match          75.0%; Score 78; DB 2; Length 227;
Best Local Similarity 94.2%; Pred. No. 3.1e-19;
Matches 81; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 18 tacactgactgcactgaatccggtgcagaacctgtgcctgtgcgaaggctctaacgtttgc 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 TATACTGACTGCACCTGAATCTGGTCAGAACCTGTGCCTGTGCGAAGGATCTAACGCTTGC 63

Qy 78 gccacgggccaacaatgcattcctggg 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 GCCCAGGGTAACAAATGCATCCTTGG 89

RESULT 15
5164304-3
; Patent No. 5164304
; APPLICANT: Johnson, Paul H.; Lazar, Jerome B.; Sohnel, Indira
; Waleh, Nahid S.
; TITLE OF INVENTION: METHOD AND VECTORS FOR STABILIZING HIRUDIN
; AND HUMAN LAMININ B1 EXPRESSION
; NUMBER OF SEQUENCES: 26
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/347,545
; FILING DATE: 04-MAY-1989
; SEQ ID NO:3:
; LENGTH: 227
5164304-3

Query Match          74.4%; Score 77.4; DB 6; Length 227;
Best Local Similarity 88.4%; Pred. No. 5.1e-19;
Matches 84; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 9 atggtttgttacactgactgcactgaatccggtcagaaacctgtgcctgtgcgaaggctct 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11 atggtgtatacactgattgcaccgaatctggtcagaacctgtgcctgtgtgaaggttcc 70

Qy 69 aacgtttgcggccagggcaacaatgcattcctggg 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 aacgtatgcggccaaggtacaagtgcattcctggg 105

Search completed: July 15, 2002, 22:31:47
Job time: 18473 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 16:34:09 ; Search time 6165.88 Seconds  
(without alignments)  
227.653 Million cell updates/sec

Title: US-10-053-641-1  
Perfect score: 104  
Sequence: 1 gatcctttaggtgtttac.....gcaacaaatgcattcctgggc 104

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Watch 0%  
Maximum Watch 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| C 1        | 33.2  | 31.9        | 535    | 10 | BI213620 RE19057.5 |
| C 2        | 31.8  | 30.6        | 204    | 10 | BI242456 RE39782.5 |
| C 3        | 31.8  | 30.6        | 207    | 9  | AA698346 HL04318.5 |
| C 4        | 31.8  | 30.6        | 239    | 10 | BI170090 RE11441.5 |
| C 5        | 31.8  | 30.6        | 244    | 10 | BI577602 RE71310.5 |
| C 6        | 31.8  | 30.6        | 251    | 10 | BF491529 AT28289.5 |
| C 7        | 31.8  | 30.6        | 264    | 10 | BI609601 RH14405.5 |
| C 8        | 31.8  | 30.6        | 274    | 10 | BI239641 RE31685.5 |
| C 9        | 31.8  | 30.6        | 277    | 10 | BI169275 RE10403.5 |
| C 10       | 31.8  | 30.6        | 278    | 10 | BI641092 SD24239.5 |
| C 11       | 31.8  | 30.6        | 287    | 10 | BI482355 RE65038.5 |
| C 12       | 31.8  | 30.6        | 306    | 10 | BI242439 RE39764.5 |
| C 13       | 31.8  | 30.6        | 306    | 10 | BI358576 RE45394.5 |
| C 14       | 31.8  | 30.6        | 328    | 10 | BI173930 RE17078.5 |
| C 15       | 31.8  | 30.6        | 332    | 9  | AA202819 LD03151.5 |
| C 16       | 31.8  | 30.6        | 343    | 10 | BI358523 RE45309.5 |
| C 17       | 31.8  | 30.6        | 348    | 10 | BI482786 RE65544.5 |

|      |      |      |     |    |          |
|------|------|------|-----|----|----------|
| C 18 | 31.8 | 30.6 | 348 | 10 | BI576148 |
| C 19 | 31.8 | 30.6 | 349 | 9  | AA696815 |
| C 20 | 31.8 | 30.6 | 351 | 9  | AI944520 |
| C 21 | 31.8 | 30.6 | 366 | 9  | AI546107 |
| C 22 | 31.8 | 30.6 | 377 | 9  | AA699240 |
| C 23 | 31.8 | 30.6 | 378 | 10 | BI165320 |
| C 24 | 31.8 | 30.6 | 378 | 10 | BI375691 |
| C 25 | 31.8 | 30.6 | 385 | 10 | BI238863 |
| C 26 | 31.8 | 30.6 | 386 | 9  | AI062229 |
| C 27 | 31.8 | 30.6 | 390 | 10 | BI572383 |
| C 28 | 31.8 | 30.6 | 391 | 9  | AA695151 |
| C 29 | 31.8 | 30.6 | 395 | 10 | BI356434 |
| C 30 | 31.8 | 30.6 | 404 | 9  | AI294830 |
| C 31 | 31.8 | 30.6 | 406 | 10 | BI237661 |
| C 32 | 31.8 | 30.6 | 408 | 10 | BI486617 |
| C 33 | 31.8 | 30.6 | 411 | 10 | BI164996 |
| C 34 | 31.8 | 30.6 | 413 | 10 | BI483817 |
| C 35 | 31.8 | 30.6 | 414 | 9  | AA567769 |
| C 36 | 31.8 | 30.6 | 422 | 10 | BI237177 |
| C 37 | 31.8 | 30.6 | 424 | 9  | AA979257 |
| C 38 | 31.8 | 30.6 | 424 | 10 | BI641601 |
| C 39 | 31.8 | 30.6 | 425 | 10 | BI162126 |
| C 40 | 31.8 | 30.6 | 427 | 10 | BI581717 |
| C 41 | 31.8 | 30.6 | 435 | 10 | BI242206 |
| C 42 | 31.8 | 30.6 | 438 | 10 | BI620889 |
| C 43 | 31.8 | 30.6 | 442 | 9  | AA201282 |
| C 44 | 31.8 | 30.6 | 443 | 9  | AA696887 |
| C 45 | 31.8 | 30.6 | 446 | 9  | AA538780 |

ALIGNMENTS

RESULT 1

BI213620/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BDGP

One Cyclotron Rd,

Lawrence Berkeley National Lab

One Cyclotron Rd,

Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>

hit genomic AE003846: arm:4 [752420..1050888]

estimated-cyto:102E2-102F6: 04/12/2001

Plate: RE190 row: E column: 9

High quality sequence stop: 474.

Location/Qualifiers

1..535

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="RE19057"

/clone\_lib="RE Drosophila melanogaster normalized Embryo

BI213620 535 bp mRNA linear EST 11-JUL-2001  
RE19057.Sprime RE Drosophila melanogaster normalized Embryo pFic-1  
Drosophila melanogaster cDNA clone RE19057 5 similar to  
ATPsyn-beta: FBan0011154 'enzyme' -AND- 'enzyme' located on: 4  
102F4-102F4;: 04/12/2001, mRNA sequence.

BI213620 1 GI:14691344  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 535)  
Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson  
J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George  
R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,  
Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,  
Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celnikier,S. and Rubin  
G.M.

BDGP/HMI RE Drosophila EST Project  
Unpublished (2001)  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
hit genomic AE003846: arm:4 [752420..1050888]  
estimated-cyto:102E2-102F6: 04/12/2001  
Plate: RE190 row: E column: 9  
High quality sequence stop: 474.  
Location/Qualifiers  
1..535  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="RE19057"  
/clone\_lib="RE Drosophila melanogaster normalized Embryo

```

pFlc-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha TONa"
/notes="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
129 a 128 c 140 g 138 t
BASE COUNT
ORIGIN

```

[illegible]

|            |   |
|------------|---|
| RESULT     | 2   |
| BI242456/c |   |
| LOCUS      | BI242456  |
| DEFINITION | 204 bp mRNA linear EST 12-JUL-2001                                  |
|            | RE39782.5 prime RE Drosophila melanogaster normalized Embryo pFlc-1 |
|            | Drosophila melanogaster cDNA clone RE39782.5, mRNA sequence.        |

| EST.       | KEYWORDS | SOURCE | ORGANISM                       | REFERENCE          | AUTHORS  |
|------------|----------|--------|--------------------------------|--------------------|--|
| fruit fly. |          |        | <i>Drosophila melanogaster</i> | 1 (bases 1 to 204) | Stapleton, M., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C. J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniaker, S. and Rubin, G. M. |

| TITLE                               | JOURNAL  | COMMENT  |
|-------------------------------------|--|--|
| BDGP/HHMI RE Drosophila EST Project | Unpublished (2001)   | Contact: Stapleton, M.<br>BDGP   |
| Lawrence Berkeley National Lab      | One Cyclotron Rd   | Berkeley, CA 94720, USA  |
| Fax: 510 486 6798                   | Email: <a href="http://www.fruitfly.org/EST">http://www.fruitfly.org/EST</a> , | <a href="mailto:est@fruitfly.berkeley.edu">est@fruitfly.berkeley.edu</a> |
| Plate: RE.397                       | row: G   | column: 10   |
| High quality                        | sequence   | stop: 198.   |

```

FEATURES
source
    .204
    Location/Qualifiers
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone="RE39782"
    /clone_lib="RE Drosophila melanogaster normalized Embryo
    pfic-1"
    /sex="male and female"
    /dev_stage="0-24 hours mixed stage embryonic"
    /lab_host="DH5-alpha TONa"
    /note="organ: embryo; Vector: pRc1; Site_1: XhoI; Site_2:
    BamHI; Library was kindly generated by Piero Carninci at
    the RIKEN. The library was normalized and excised using
    Cre recombinase. Plasmid cDNA library."
    50 a 44 c 56 q 54 t
BASE COUNT

```

BASE COUNT  
ORIGIN

|    | Query Match           | 30.6%   | Score 31.8;    | DB 10;    | Length 204; |
|----|-----------------------|---|----------------|-----------|-------------|
|    | Best Local Similarity | 67.2%;  | Pred. No. 2.7; |           |             |
|    | Matches 45;           | Conservative 0;   | Mismatches 22; | Indels 0; | Gaps 0;     |
| Qy | 24                    | gactgcactgaatccgggtcagaaacctgtgctgtgcgaaggctctctaacgitttgcgccag | 83             |           |             |
|    |                       |   |                |           |             |
|    |                       |   |                |           |             |
|    |                       |   |                |           |             |
|    |                       |   |                |           |             |
| Db | 153                   | GACTGCGACCAATTACGGCCACAATCTTTCCATTTCGGCAGTGCAGCCCTTGCGACCTT     | 94             |           |             |
|    |                       |   |                |           |             |
|    |                       |   |                |           |             |
| Qy | 84                    | ggcaaca   | 90             |           |             |
|    |                       |   |                |           |             |
| Db | 93                    | GGCAGCA   | 87             |           |             |

|            |                         |  |
|------------|-------------------------|--|
| RESULT     | 3                       |  |
| AA698346/c |                         |  |
| LOCUS      |                         |  |
| DEFINITION | AA698346                | 207 bp mRNA linear EST 19-APR-2001         |
|            | HL04318.5prime          | HL Drosophila melanogaster head BlueScript |
|            | Drosophila melanogaster | cDNA clone HL04318 5prime, mRNA sequence.  |
| ACCESSION  | AA698346                |  |
| VERSION    | AA698346.1              | GI:2701275                                 |
| KEYWORDS   | EST.                    |  |
| SOURCE     | fruit fly.              |  |
| ORGANISM   | Drosophila melanogaster |  |

|           |  |
|-----------|--|
| REFERENCE | muscomorpha, ephyridae; drosoptillidae; drosoptilla.<br>1 (bases 1 to 207)                           |
| AUTHORS   | Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,<br>Lewis, S. and Rubin, G.M. |
| TITLE     | BDGP/HIMI Drosoptilla EST Project  |
| JOURNAL   | Unpublished (2001)   |
| COMMENT   | Contact: Stapleton, M.<br>BDGP   |

Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
Place: 43 row: B column: 6  
High quality sequence stop: 161.

```

FEATURES
source
high quality sequence stop: 101.
location/Qualifiers
1..207
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="HL04318"
/clone_lib="HL Drosophila melanogaster head BlueScript"
/sex="male and female"
/dev_stage="adult"
/lab_host="SOLR"
/note="Organ: head-brain & sensory organ; Vector:
BlueScript SK; Site:1: EcoRI; Site:2: XhoI; Constructed
using Stratagene ZAP-cDNA Synthesis kit; Oligo dT-primed
and directionally cloned at EcoRI and XhoI in Bluescript
SK(+)-4318"

```

|                       |      |         |        |                |                   |
|-----------------------|------|---------|--------|----------------|-------------------|
| BASE COUNT            | 49 a | 5N(+/+) | 48 c   | 54 g           | 56 t              |
| ORIGIN                |      |         |        |                |                   |
| Query Match           |      |         | 30.6%  | Score 31.8;    | DB 9; Length 207; |
| Best local Similarity |      |         | 67.2%; | Pred. No. 2.7; |                   |

|    |  |    |
|----|--|----|
| 24 | gactgcaactgaatccggtcagaacctgtgcctgtgcgaagctctataacgtttgcggccag | 83 |
| Oy |  |    |
| 12 | gactgcaccaattacggcgacaattttccattttgCGCAGCTGcAGCCTTTGcAGCCTT    | 83 |
| Db |  |    |
| 84 | ggcaaca  | 90 |
| Oy |  |    |
| 82 | ggcagca  | 76 |
| Db |  |    |

| RESULT      | 4 |        |      |        |                 |
|-------------|---|--------|------|--------|-----------------|
| BIL170090/c |   |        |      |        |                 |
| LOCUS       |   |        |      |        |                 |
| BIL170090   |   | 239 bp | mrna | linear | EST 09-JUL-2001 |

```

DEFINITION RE11441.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
Drosophila melanogaster cDNA clone RE11441 5, mRNA sequence.
ACCESSION BI170090
VERSION BI170090.1 GI:14635897
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 239)
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin
G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: RE.114 row: D column: 5
High quality sequence stop: 232.
Location/Qualifiers
FEATURES
source
1..239
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RE11441"
/clone="RE11441"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/notes="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT 57 a 55 c 63 g 64 t
ORIGIN
Query Match 30.6%; Score 31.8; DB 10; Length 239;
Best Local Similarity 67.2%; Pred. NO. 2.8;
Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 24 gactgcactgaatccggtcagaaacctgtgctgtgctgaaggctctaaagctttgaggccag 83
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 153 GACTGCACCAATACGGCCCAATCTTTCATTTGCGGACGCTGCGAGCCTTTGCGAGCCTT 94
QY 84 ggcacaa 90
Db 93 GGCAGCA 87
RESULT 5
LOCUS BI577602
DEFINITION RE17310.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
Drosophila melanogaster cDNA clone RE17310 5 similar to
ATPsyn-beta: FBan011154 GO: [hydrogen-transporting ATP synthase
(GO:0003936); hydrogen-transporting ATP synthase, catalytic core
CF(1) (GO:0005754); hydrogen-transporting ATP synthase (GO:0003936
)] located on: 4 102F4-102F4.; 05/21/2001, mRNA sequence.
ACCESSION BI577602
VERSION BI577602.1 GI:15469024
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 244)
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin
G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: RE.713 row: A column: 10
High quality sequence stop: 243.
Location/Qualifiers
FEATURES
source
1..244
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RE17310"
/clone="RE17310"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/notes="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT 57 a 55 c 66 g 66 t
ORIGIN
Query Match 30.6%; Score 31.8; DB 10; Length 244;
Best Local Similarity 67.2%; Pred. NO. 2.8;
Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 24 gactgcactgaatccggtcagaaacctgtgctgtgctgaaggctctaaagctttgaggccag 83
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 157 GACTGCACCAATACGGCCCAATCTTTCATTTGCGGACGCTGCGAGCCTTTGCGAGCCTT 98
QY 84 ggcacaa 90
Db 97 GGCAGCA 91
RESULT 6
LOCUS BF491529
DEFINITION AT28289.5prime AT Drosophila melanogaster adult testes pOTB7
Drosophila melanogaster cDNA clone AT28289 5 similar to
ATPsyn-beta: FBan011154 'enzyme' -AND- 'enzyme' located on: 4
102F4-102F4.; 04/09/2001, mRNA sequence.
ACCESSION BF491529
VERSION BF491529.2 GI:13695056
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 251)
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman
B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan
D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,
Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J.,
Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E.,
Celniker, S. and Rubin, G.M.
TITLE BDGP/HMI AT Drosophila EST Project

```

1



•





Search completed: July 15, 2002, 21:48:47  
Job time: 18878 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:38:32 ; Search time 2368.24 Seconds

(without alignments)  
945.487 Million cell updates/sec

Title: US-10-053-641-2

Perfect score: 107

Sequence: 1 ctctagagccagatgcat.....cagttaaacacataaag 107

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_lov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
|------------|-------|-------------|--------|----|-------------|

RESULT 1  
SYNHIR/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

CDS

SYNHIR 226 bp DNA linear SYN 27-APR-1993  
Synthetic hirudin gene, complete cds.

M26762.1 GI:208478

hirudin; proteinase inhibitor; thrombin inhibitor.

Synthetic DNA.

artificial sequence.

1 (bases 1 to 226)

Bergmann, C., Dodt, J., Koehler, S., Fink, E. and Gassen, H.G.

Chemical synthesis and expression of a gene coding for hirudin, the

thrombin-specific inhibitor from the leech Hirudo medicinalis

Biol. Chem. Hoppe-Seyler 367, 731-740 (1986)

87026239

Location/Qualifiers

1..226

/organism="synthetic construct"

/db\_xref="taxon:32630"

22..222

/note="hirudin"

/codon\_start=1

ALIGNMENTS

```

/transl_table=11
/protein_id="AAA72772.1"
/db_xref="GI:208479"
/translation="MVVYDCTESGNLCLCEGSNNVCGGCKNLILSGDEKNCQVTE
GTPKQSHNDGFEEIPEYIQ"
63 a 56 c 61 g 46 t
BASE COUNT
ORIGIN

```

|                       |                 |                    |           |             |
|-----------------------|-----------------|--------------------|-----------|-------------|
| Query Match           | 96.3%           | Score 103;         | DB 12;    | Length 236; |
| Best Local Similarity | 100.0%;         | Pred. No. 8.5e-23; |           |             |
| Matches 103;          | Conservative 0; | Mismatches 0;      | Indels 0; | Gaps 0;     |

|    |     |   |    |
|----|-----|---|----|
| Qy | 5   | agagcccgagatgattgttgccttggcgcgcaaaagttagagccttcgcacaggcacag | 64 |
|    |     |   |    |
|    |     |   |    |
|    |     |   |    |
|    |     |   |    |
| Db | 120 | AGAGCCCGAGGATGATTTGTTGCCCTGGCGCAACGTTAGAGCCTTCGCACAGGCACAG  | 61 |

|    |    |   |     |
|----|----|---|-----|
| Qy | 65 | gtctgaccggattcagtcgagtcagtgtaaacacaccataaag     | 107 |
|    |    |   |     |
|    |    |   |     |
|    |    |   |     |
| Db | 60 | GTCTGACCGGATTTCAGTCGATGCAGTCAGTGTAAACAACCATAAAG | 18  |

|            |  |
|------------|--|
| RESULT     | 2  |
| LOCUS      | I41183/c   |
| DEFINITION | I41183   |
| ACCESSION  | Sequence 3 from patent US 5624822.                   |
| KEYWORDS   | I41183<br>I41183.1 GI:2081773                        |
| SOURCE     | Unknown.   |
| ORGANISM   | Unknown.<br>Unclassified.                            |
| REFERENCE  | 1 (bases 1 to 129)                                   |
| AUTHORS    | Koerwer, W.  |
| TITLE      | Hirudin fusion proteins and preparation of hirudin   |
| JOURNAL    | Patent: US 5624822-A 3 29-APR-1997;                  |
| FEATURES   | Location/Qualifiers<br>1..129<br>/organism="unknown" |
| source     | 36 a 35 c 29 g 29 t                                  |
| BASE COUNT |  |
| ORIGIN     |  |

|    |                       |  |                    |       |             |         |
|----|-----------------------|--|--------------------|-------|-------------|---------|
|    | Query Match           | 92.5%;   | Score 99;          | DB 6; | Length 129; |         |
|    | Best Local Similarity | 100.0%;  | Pred. No. 1.7e-21; |       |             |         |
|    | Matches               | 99;  | Conservative       | 0;    | Mismatches  | 0; Gaps |
|    |                       |  |                    |       |             | 0;      |
| QY | 5                     | agagccacgagatgcatttggccctggcgcaaacgttagcgttcacctgcacaggcacag | 64                 |       |             |         |
|    |                       |  |                    |       |             |         |
| Db | 129                   | AGAGCCACGAGATGCATTGTGGCCCTGGCGCAACGTTAGAGCCTTCGCACAGGCCACAG  | 70                 |       |             |         |
|    |                       |  |                    |       |             |         |
| QY | 65                    | gtttcacccgatttcagtgcagtcagtgtaataacaaccat                    | 103                |       |             |         |
|    |                       |  |                    |       |             |         |
| Db | 69                    | GTTCTGACCGGATTTCAGTGCAGTCAGTGTAACAACCAT                      | 31                 |       |             |         |
|    |                       |  |                    |       |             |         |

|            |  |        |        |
|------------|--|--------|--------|
| RESULT     | 3  |        |        |
| I41187/c   |  |        |        |
| LOCUS      | I41187   | 235 bp | DNA    |
| DEFINITION | Sequence 7 from patent US 5624822.                 |        | linear |
| ACCESSION  | I41187   |        |        |
| VERSION    | I41187.1   |        |        |
| KEYWORDS   | GI:2081777   |        |        |
| SOURCE     | .  |        |        |
| ORGANISM   | Unknown.   |        |        |
| REFERENCE  | Unknown.   |        |        |
| AUTHORS    | Unclassified.                                      |        |        |
| TITLE      | 1 (bases 1 to 235)                                 |        |        |
| JOURNAL    | Koerwer, W.  |        |        |
| FEATURES   | Hirudin fusion proteins and preparation of hirudin |        |        |
| source     | Patent: US 5624822-A 7 29-APR-1997;                |        |        |
|            | Location/Qualifiers                                |        |        |
|            | 1..235   |        |        |
|            | /organism="unknown"                                |        |        |

•

```

BASE COUNT      72 a    63 c    57 g    43 t
ORIGIN
Query Match          92.5%; Score 99; DB 6; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.7e-21;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   5 agagccaggatgcattgttgccctggcgcgcaaacgttagagccttcgcacaggcacag 64
      |||
Db   129 AGAGCCACGAGTGCATTTGTGGCTCGGCCGCAACACTTAGAGCCTTCGCACAGGCACAG 70

Qy   65 gtctgaccggattcagtcgagtcagtcagtgtaaacaacct 103
      |||||
Db   69 GTTCTGACCGGATTTCAGTGCGAGTCAGTGTAACAACCAT 31

RESULT  4
LOCUS       I41188                               235 bp    DNA        linear   PAT 13-MAY-1997
DEFINITION Sequence 8 from patent US 5624822.
ACCESSION   I41188
VERSION     I41188.1 GI:2081778
KEYWORDS    .
SOURCE      Unknown.
            Unclassified.
            Organism      Unclassified.
REFERENCE   1 (bases 1 to 235)
AUTHORS    Koerwer,W.
TITLE       Hirudin fusion proteins and preparation of hirudin
JOURNAL    Patent: US 5624822-A 8 29-APR-1997;
FEATURES    Location/Qualifiers
            1..235
            /organism="unknown"

```

|    | Query Match           | 92.5%   | Score 99;    | DB 6;    | Length 235; |
|----|-----------------------|---|--------------|----------|-------------|
|    | Best Local Similarity | 100.0%;   | Pred. No.    | 1.7e-21; |             |
|    | Matches               | 99;   | Conservative | 0;       | Mismatches  |
|    |                       |   |              |          | Indels      |
|    |                       |   |              |          | Gaps        |
|    |                       |   |              |          | 0;          |
| Qy | 5                     | agagccacgagtcatgttgccttgggcgcgaacatttagacccttcgcacaggcacag  | 64           |          |             |
|    |                       |   |              |          |             |
|    |                       |   |              |          |             |
|    |                       |   |              |          |             |
|    |                       |   |              |          |             |
| Db | 111                   | AGAGCCACGATGCATTGTGGTCCCTCGGCCGAAGCATTAGAGCCTTCGCACAGGCACAG | 170          |          |             |
|    |                       |   |              |          |             |
|    |                       |   |              |          |             |
|    |                       |   |              |          |             |
|    |                       |   |              |          |             |
| Qy | 65                    | gtttcgacctgattcaatgcagtcagtcagtgtaaacacaacct                | 103          |          |             |
|    |                       |   |              |          |             |
|    |                       |   |              |          |             |
|    |                       |   |              |          |             |
| Db | 171                   | GTCTGTACC GGATTCAGTGCACTCATGTGTAACAACCAT                    | 209          |          |             |
|    |                       |   |              |          |             |
|    |                       |   |              |          |             |
|    |                       |   |              |          |             |
|    |                       |   |              |          |             |

|             |  |            |                   |
|-------------|--|------------|-------------------|
| RESULT      | 5  |            |                   |
| 141189/c    |  |            |                   |
| LOCUS       | 141189   | 726 bp     | DNA               |
| DEFINITION  | Sequence 9 from patent US 5624822.                 |            | linear            |
| ACCESSION   | 141189   |            |                   |
| VERSION     | 141189.1   | GI:2081779 |                   |
| KEYWORDS    | .  |            |                   |
| SOURCE      | Unknown.   |            |                   |
| ORGANISM    | Unknown.   |            |                   |
| REFERENCE   | Unclassified.                                      |            |                   |
| AUTHORS     | 1 (bases 1 to 726)                                 |            |                   |
| TITLE       | Koerwer/W.   |            |                   |
| JOURNAL     | Hirudin fusion proteins and preparation of hirudin |            |                   |
| FEATURES    | Patent: US 5624822-A 9 29-APR-1997;                |            |                   |
| source      | Location/Qualifiers                                |            |                   |
|             | 1..726   |            |                   |
|             | /organism="unknown"                                |            |                   |
| BASE COUNT  | 281 a  | 171 c      | 148 t             |
| ORIGIN      |  |            |                   |
| Query Match | 92.5%  | Score 99:  | DB 6: Length 726: |













```

XX PI Korwer W;
XX FT WPI; 1991-194236/27.
XX DR P-PSDB; AAR12751.
XX XX
XX XX Hirudin peptide prodn. by cleaving new fusion peptide - of
XX PT hirudin and protein A, expressed in high yield and stable,
XX PT soluble form by transformed E.coli
XX XX
XX PS Example 1; Page 6-7; 9pp; German.
XX CC This sequence is an example of a fusion construct for expression of
XX CC the fusion peptide of the invention. The Met residue in the linker
XX CC oligopeptide allows cleavage by CNBR to release two fragments which
XX CC can be easily separated by IGF affinity chromatography. (The protein
XX CC A component binds to IGF sepharose). Increased yields of hirudin are
XX CC obtained using this fusion construct.
XX SQ Sequence 726 BP; 280 A; 171 C; 126 G; 149 T; 0 other;

Query Match 92.5%; Score 99; DB 12; Length 726;
Best Local Similarity 100.0%; Pred. No. 1.4e-24;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 agagccaggatgattgttgcctgcccgcgaacgttagagccttcgcacaggcacag 64
Db 621 AGAGCCAGGATGATTTGTCCTGCGCGCAACGTTAGAGCCTTCGCACAGGCACAG 562

Qy 65 gttctaccggatcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 103
Db 561 GTTCTGACCGGATTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 523

RESULT 2
AAQ93244/c
ID AAQ93244 standard; DNA; 468 BP.
XX AC AAQ93244;
XX DT 01-NOV-1995 (first entry)
XX XX Fusion construct of glucoamylase-hirudin DNA.
XX KW expression cassette; recombinant protein; production; Hansenula;
XX KW yeast; leader sequence; adaptor; alpha helix; glucoamylase;
XX KW secretion; processing; thrombin inhibitor; hirudin; ds.
XX OS Synthetic.
XX PH Key
XX FT primer_bind Location/Qualifiers
XX FT complement (1..33)
XX FT /tag= a
XX FT /note= "primer AAQ85840 binding site"
XX FT 13..444
XX FT /tag= b
XX FT /product= glucoamylase-hirudin fusion protein
XX FT 199..223
XX FT /tag= c
XX FT /note= "primer AAQ85841 binding site"
XX FT complement (220..267)
XX FT /tag= d
XX FT /note= "primer AAQ85842 binding site"
XX FT 13..228
XX FT /tag= e
XX FT /note= "encodes amino acids 1-72 of glucoamylase"
XX FT 241..246
XX FT /tag= f
XX FT /note= "encodes processor signal"
XX FT 247..444
XX FT /tag= g
XX FT /note= "encodes hirudin-h120"

```

```

FT primer_bind 0..468
FT /tag= h
FT /note= "primer AAQ85843 binding site"
XX PN DE4329969-A.
XX XX
XX PD 09-MAR-1995.
XX PF 04-SEP-1993; 93DE-4329969.
XX PR 04-SEP-1993; 93DE-4329969.
XX PA (BADI ) BASF AG.
XX PA (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
XX PI Bollschweiler C, Janowicz ZA, Piontek M, Schweden J;
XX PI Strasserawm, Weydemann U;
XX DR WPI; 1995-107810/15.
XX DR P-PSDB; AAR76951.
XX XX Recombinant protein prodn. in Hansenula yeast - transformed with
XX PT expression cassette contg. leader, adaptor, processing signal and
XX PT gene, provides efficient secretion and correct processing
XX PS Example 1; Fig 1; 10pp; German.
XX XX AAQ93244 is a DNA fusion product resulting from ligation of an
XX CC EcoRI-PvuI fragment (see AAQ85840-41) and a PvuI-SalI fragment, encoding
XX CC a leader sequence, an adaptor (see AAR71472), amino acids 23-72 of GAM
XX CC (glucamylase from Schwanniomycetes occidentalis, plus -His-Pro-Leu-Gln at
XX CC the C-terminal end), a processor peptide (Lys-Arg) and a structural gene,
XX CC encoding hirudin. AAQ93244 is an example of an expression vector insert
XX CC of the general formula: L-A-P-Gene (L = leader sequence; A = adaptor;
XX CC P = processor; Gene = structural gene). The cassettes ensure efficient
XX CC secretion and correct processing of heterologous structural genes in
XX CC yeast of the genus Hansenula, and so provides high yields of mature
XX CC proteins and facilitates subsequent purification.
XX SQ Sequence 468 BP; 126 A; 110 C; 112 G; 120 T; 0 other;

Query Match 92.0%; Score 98.4; DB 16; Length 468;
Best Local Similarity 99.0%; Pred. No. 2e-24;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ctctagagcccaggatgcattgttgcctgcccgcgaacgttagagccttcgcacaggc 60
Db 346 CTTTAGAGCCCGAGGATGATTTGTCCTGCGCGCAACGTTAGAGCCTTCGCACAGGC 287

Qy 61 acaggttctgaccggatcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 100
Db 286 ACAGGTTCTGACCGGATTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 247

RESULT 3
AAQ60746/c
ID AAQ60746 standard; DNA; 224 BP.
XX AC AAQ60746;
XX XX
XX DT 01-JAN-1980 (first entry)
XX XX DNA encoding hirudin-like polypeptide.
XX DE Hirudin; thrombin-antagonist; anticoagulant; ds.
XX KW Synthetic.
XX OS DE3445517-A.
XX PN 19-JUN-1986.
XX PD
XX XX

```



96 AGAACCCAGGATGCATTTGTACCCCTGACCCGCAACGTTAGAACCTTCGCACAGGCACAG 37

117 GTTCTGACCAAGATTCCGGTGCAGTCGGGTAAACAAC 82

RESULT 8



















```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-262-384A-6

Query Match 75.7%; Score 81; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 2.8e-20;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 gttgcccctggccgcaaacgttagagccttcgcacagggcacaggttctgacccggttcagt 82
Db 1 GTTGCCCTGGCCGCAACGTTAGAGCCTTCGCACAGGCACAGGTTCTGACCGGATTTCAGT 60

Qy 83 gcagtcagtgtaaacacacat 103
Db 61 GCAGTCAGTGTAACAACCAT 81

RESULT 11
US-08-757-439-1
; Sequence 1, Application US/08757439
; Patent No. 5866371
; GENERAL INFORMATION:
; APPLICANT: BADZIONG, Werner
; APPLICANT: HABERMANN, Paul
; APPLICANT: MOELLER, Joerg
; APPLICANT: ARETZ, Werner
; TITLE OF INVENTION: PROMOTER SYSTEM FOR USING THE YEAST ADH II
; TITLE OF INVENTION: PROMOTER SYSTEM FOR THE PRODUCTION OF HETEROLOGOUS
; TITLE OF INVENTION: PROTEINS IN HIGH YIELDS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,439
; FILING DATE: 27-NOV-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19544233.4
; FILING DATE: 28-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: SANDERCOCK, Colin G.
; REGISTRATION NUMBER: 31,298
; REFERENCE/DOCKET NUMBER: 18748/303/HOCE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-757-439-1
Query Match 74.4%; Score 79.6; DB 2; Length 8491;
Best Local Similarity 90.4%; Pred. No. 3.7e-19;
Matches 85; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 6 gagcccgagtgatcttggccctggccgcaaacgttagagccttcgcacagggcacag 65
Db 6893 GATCCAAGGATGATTTGTTACCTGGCCGCAACGTTAGATCTCTTCGCACAGGCACAGG 6952

Qy 66 ttctgaccggattcagtgcaagtcagtgtaaacaa 99
Db 6953 TTCTGACCAGATTCAAGTCAGTCAGTCAGTATACGTAA 6986

RESULT 12
5180668-10/c
; Patent No. 5180668
; APPLICANT: CRAUSE, PETER; HABERMANN, PAUL; TRIPIER, DOMINIQUE
; TITLE OF INVENTION: HIRUDIN DERIVATIVE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295,422
; FILING DATE: 10-JAN-1989
; SEQ ID NO: 10;
; LENGTH: 212
; 5180668-10

Query Match 74.2%; Score 79.4; DB 6; Length 212;
Best Local Similarity 93.3%; Pred. No. 1.3e-19;
Matches 83; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 6 gagcccgagtgatcttggccctggccgcaaacgttagagccttcgcacagggcacag 65
Db 100 GATCCAAGGATGATTTGTTACCTGGCCGCAACGTTAGATCTCTTCGCACAGGCACAGG 41

Qy 66 ttctgaccggattcagtgcaagtcagtgta 94
Db 40 TTCTGACCAGATTCAAGTCAGTCAGTCAGTATA 12

RESULT 13
US-07-982-064-8/c
; Sequence 8, Application US/07982064
; Patent No. 5918895
; GENERAL INFORMATION:
; APPLICANT: Schmid, G.; Habermann, P.
; TITLE OF INVENTION: Secretion of Hirudin Derivatives
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Collard, Roe & Galgano, P.C.
; STREET: 1077 No. 5918895thern Boulevard
; CITY: Roslyn
; STATE: New York
; COUNTRY: USA
; ZIP: 11576
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/982,064
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/672,791
; FILING DATE: 21-MAR-1991
; APPLICATION NUMBER: GR 40 09 268.2
; FILING DATE: 22 MAR 1990
; ATTORNEY/AGENT INFORMATION:
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:48:47 ; Search time 6165.88 Seconds  
(without alignments)  
234.220 Million cell updates/sec

Title: US-10-053-641-2

Perfect score: 107

Sequence: 1 cttctagagccaggatgcat.....cagtgtaaacacaccataaag 107

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

|     |              |
|-----|--------------|
| 1:  | em_estba:*   |
| 2:  | em_esthum:*  |
| 3:  | em_estin:*   |
| 4:  | em_estmu:*   |
| 5:  | em_estov:*   |
| 6:  | em_estpl:*   |
| 7:  | em_estro:*   |
| 8:  | em_htc:*     |
| 9:  | gb_est1:*    |
| 10: | gb_est2:*    |
| 11: | gb_htc:*     |
| 12: | gb_gss:*     |
| 13: | em_gss_hum:* |
| 14: | em_gss_inv:* |
| 15: | em_gss_pln:* |
| 16: | em_gss_vrt:* |

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 33.2  | 31.0        | 535    | 10    | BI213620    |
| 2          | 31.8  | 29.7        | 204    | 10    | BI242456    |
| 3          | 31.8  | 29.7        | 207    | 9     | AA698346    |
| 4          | 31.8  | 29.7        | 239    | 10    | BI170090    |
| 5          | 31.8  | 29.7        | 244    | 10    | BI577602    |
| 6          | 31.8  | 29.7        | 251    | 10    | BF491529    |
| 7          | 31.8  | 29.7        | 264    | 10    | BI609601    |
| 8          | 31.8  | 29.7        | 274    | 10    | BI239641    |
| 9          | 31.8  | 29.7        | 277    | 10    | BI169275    |
| 10         | 31.8  | 29.7        | 278    | 10    | BI641092    |
| 11         | 31.8  | 29.7        | 287    | 10    | BI482355    |
| 12         | 31.8  | 29.7        | 306    | 10    | BI242439    |
| 13         | 31.8  | 29.7        | 306    | 10    | BI358576    |
| 14         | 31.8  | 29.7        | 328    | 10    | BI173930    |
| 15         | 31.8  | 29.7        | 332    | 9     | AA02819     |
| 16         | 31.8  | 29.7        | 343    | 10    | BI358523    |
| 17         | 31.8  | 29.7        | 348    | 10    | BI482786    |

|    |      |      |     |    |          |
|----|------|------|-----|----|----------|
| 18 | 31.8 | 29.7 | 348 | 10 | BI576148 |
| 19 | 31.8 | 29.7 | 349 | 9  | AA696815 |
| 20 | 31.8 | 29.7 | 351 | 9  | AI944520 |
| 21 | 31.8 | 29.7 | 366 | 9  | AI546107 |
| 22 | 31.8 | 29.7 | 377 | 9  | AA699240 |
| 23 | 31.8 | 29.7 | 378 | 10 | BI165320 |
| 24 | 31.8 | 29.7 | 378 | 10 | BI375691 |
| 25 | 31.8 | 29.7 | 385 | 10 | BI238863 |
| 26 | 31.8 | 29.7 | 386 | 9  | AI062229 |
| 27 | 31.8 | 29.7 | 390 | 10 | BI572383 |
| 28 | 31.8 | 29.7 | 391 | 9  | AA695151 |
| 29 | 31.8 | 29.7 | 395 | 10 | BI356434 |
| 30 | 31.8 | 29.7 | 404 | 9  | AI294830 |
| 31 | 31.8 | 29.7 | 406 | 10 | BI237661 |
| 32 | 31.8 | 29.7 | 408 | 10 | BI486617 |
| 33 | 31.8 | 29.7 | 411 | 10 | BI164996 |
| 34 | 31.8 | 29.7 | 413 | 10 | BI483817 |
| 35 | 31.8 | 29.7 | 414 | 9  | AA567769 |
| 36 | 31.8 | 29.7 | 422 | 10 | BI237177 |
| 37 | 31.8 | 29.7 | 424 | 9  | AA979257 |
| 38 | 31.8 | 29.7 | 424 | 10 | BI641601 |
| 39 | 31.8 | 29.7 | 425 | 10 | BI162126 |
| 40 | 31.8 | 29.7 | 427 | 10 | BI581717 |
| 41 | 31.8 | 29.7 | 435 | 10 | BI242206 |
| 42 | 31.8 | 29.7 | 438 | 10 | BI620889 |
| 43 | 31.8 | 29.7 | 442 | 9  | AA201282 |
| 44 | 31.8 | 29.7 | 443 | 9  | AA696887 |
| 45 | 31.8 | 29.7 | 446 | 9  | AA538780 |

## ALIGNMENTS

### RESULT 1

#### BI213620

#### LOCUS

#### DEFINITION

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### COMMENT

#### FEATURES

#### source

BI213620 535 bp mRNA linear EST 11-JUL-2001  
 RE19057.5prime RE Drosophila melanogaster normalized Embryo pFlc-1  
 Drosophila melanogaster cDNA clone RE19057 5 similar to  
 ATPsyn-beta: FBan0011154 'enzyme' -AND- 'enzyme' located on: 4  
 102F4-102F4: 04/12/2001, mRNA sequence.

BI213620  
 BI213620.1 GI:14691344

EST.  
 fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 535)

Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson

,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George

,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,

Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,

Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celnikier,S. and Rubin

,G.W.

BDGP/HMI RE Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)

hit genomic AE003846: arm:4 [752420.1050888]

estimated-cyto:102E2-102F6: 04/12/2001

Plate: RE.190 row: E column: 9

High quality sequence stop: 474.

Location/Qualifiers

1..535

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="RE19057"

/clone\_lib="RE Drosophila melanogaster normalized Embryo







JOURNAL  
COMMENT

Unpublished (2000)  
On Dec 6, 2000 this sequence version replaced gi:11574830.  
Contact: Stapleton, M.  
RNGP

**BUDF**  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
Plate: AT:282 row: H column: 5  
High quality sequence stop: 182.

| FEATURES           | SOURCE |
|--------------------|--------|
| 1. <i>General</i>  |        |
| 2. <i>Specific</i> |        |
| 3. <i>Other</i>    |        |
| 4. <i>Other</i>    |        |
| 5. <i>Other</i>    |        |
| 6. <i>Other</i>    |        |
| 7. <i>Other</i>    |        |
| 8. <i>Other</i>    |        |
| 9. <i>Other</i>    |        |
| 10. <i>Other</i>   |        |
| 11. <i>Other</i>   |        |
| 12. <i>Other</i>   |        |
| 13. <i>Other</i>   |        |
| 14. <i>Other</i>   |        |
| 15. <i>Other</i>   |        |
| 16. <i>Other</i>   |        |
| 17. <i>Other</i>   |        |
| 18. <i>Other</i>   |        |
| 19. <i>Other</i>   |        |
| 20. <i>Other</i>   |        |
| 21. <i>Other</i>   |        |
| 22. <i>Other</i>   |        |
| 23. <i>Other</i>   |        |
| 24. <i>Other</i>   |        |
| 25. <i>Other</i>   |        |
| 26. <i>Other</i>   |        |
| 27. <i>Other</i>   |        |
| 28. <i>Other</i>   |        |
| 29. <i>Other</i>   |        |
| 30. <i>Other</i>   |        |
| 31. <i>Other</i>   |        |
| 32. <i>Other</i>   |        |
| 33. <i>Other</i>   |        |
| 34. <i>Other</i>   |        |
| 35. <i>Other</i>   |        |
| 36. <i>Other</i>   |        |
| 37. <i>Other</i>   |        |
| 38. <i>Other</i>   |        |
| 39. <i>Other</i>   |        |
| 40. <i>Other</i>   |        |
| 41. <i>Other</i>   |        |
| 42. <i>Other</i>   |        |
| 43. <i>Other</i>   |        |
| 44. <i>Other</i>   |        |
| 45. <i>Other</i>   |        |
| 46. <i>Other</i>   |        |
| 47. <i>Other</i>   |        |
| 48. <i>Other</i>   |        |
| 49. <i>Other</i>   |        |
| 50. <i>Other</i>   |        |
| 51. <i>Other</i>   |        |
| 52. <i>Other</i>   |        |
| 53. <i>Other</i>   |        |
| 54. <i>Other</i>   |        |
| 55. <i>Other</i>   |        |
| 56. <i>Other</i>   |        |
| 57. <i>Other</i>   |        |
| 58. <i>Other</i>   |        |
| 59. <i>Other</i>   |        |
| 60. <i>Other</i>   |        |
| 61. <i>Other</i>   |        |
| 62. <i>Other</i>   |        |
| 63. <i>Other</i>   |        |
| 64. <i>Other</i>   |        |
| 65. <i>Other</i>   |        |
| 66. <i>Other</i>   |        |
| 67. <i>Other</i>   |        |
| 68. <i>Other</i>   |        |
| 69. <i>Other</i>   |        |
| 70. <i>Other</i>   |        |
| 71. <i>Other</i>   |        |
| 72. <i>Other</i>   |        |
| 73. <i>Other</i>   |        |
| 74. <i>Other</i>   |        |
| 75. <i>Other</i>   |        |
| 76. <i>Other</i>   |        |
| 77. <i>Other</i>   |        |
| 78. <i>Other</i>   |        |
| 79. <i>Other</i>   |        |
| 80. <i>Other</i>   |        |
| 81. <i>Other</i>   |        |
| 82. <i>Other</i>   |        |
| 83. <i>Other</i>   |        |
| 84. <i>Other</i>   |        |
| 85. <i>Other</i>   |        |
| 86. <i>Other</i>   |        |
| 87. <i>Other</i>   |        |
| 88. <i>Other</i>   |        |
| 89. <i>Other</i>   |        |
| 90. <i>Other</i>   |        |
| 91. <i>Other</i>   |        |
| 92. <i>Other</i>   |        |
| 93. <i>Other</i>   |        |
| 94. <i>Other</i>   |        |
| 95. <i>Other</i>   |        |
| 96. <i>Other</i>   |        |
| 97. <i>Other</i>   |        |
| 98. <i>Other</i>   |        |
| 99. <i>Other</i>   |        |
| 100. <i>Other</i>  |        |

```

1. .251
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="AT28289"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
AT.121-AT.319: DH5-alpha Tona"
/note="Organ: ADULT testes; Vector: pOTB7; Site:1: ECORI;
Site:2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."
58 a 61 c 68 g 64 t

```

Query Match  
Best Local s

|                       |              |                |                |             |
|-----------------------|--------------|----------------|----------------|-------------|
| Query Match           | 29.7%        | Score 31.8;    | DB 10;         | Length 251; |
| Best Local Similarity | 67.2%;       | Pred. No. 3.8; |                |             |
| Matches 45;           | Conservative | 0;             | Mismatches 22; | Indels 0;   |
|                       |              |                | Gaps           | 0;          |

Qy 22 ttgttcctggcgcaaacgcttagacgttgcacaggcacagattctgaccgattcag 81  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 78 TGTCTCCAGGCTCAAAAGGCTCAGCTGCCGAATCGAAGAATTGTGGCCCTAATTGG 137

Qy 82 tgcagtc 88

Qy 82 tgcagtc 88  
|||||  
Db 138 TGCAGTC 144

RESULT  
BT609601

| LOCUS    | DEFINITION   | 264 bp | mRNA | linear | EST 07-SEP-2001 |
|----------|--|--------|------|--------|-----------------|
| BT609601 | RH14405:5prime RH <i>Drosophila</i> melanogaster cDNA clone RH14405 5 similar to |        |      |        |                 |
|          | ATPSyn-beta: Fly001154 GO:hydrogen-transporting ATP synthase                     |        |      |        |                 |
|          | CF1 (GO:0003936): hydrogen-transporting ATP synthase, catalytic core             |        |      |        |                 |
|          | CF1 (GO:0005754): hydrogen-transporting ATP synthase (GO:0003936                 |        |      |        |                 |
|          | ) located on 4 102F4-102F4; 08/17/2001, mRNA sequence.                           |        |      |        |                 |

ACCESSION  
VERSION

EST. SOURCE ORGANISM

fruit fly.  
Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 264)

REFERENCE

## AUTHORS

R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C. J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin, G. M.

**TITLE**  
**JOURNAL**

COMMENT  
unpublished (2001)  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
 Plate: RH.144 row: A column: 5  
 High quality sequence stop: 246.

| FEATURES           | SOURCE |
|--------------------|--------|
| 1. <i>General</i>  |        |
| 2. <i>Specific</i> |        |
| 3. <i>Other</i>    |        |

```

i. .264
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RH14405"
/clone_lib="RH Drosophila melanogaster normalized Head
pPIC-1"
/sex="male and female"
/dev_stage="Adult"
/lab_host="DHS-alpha Tona"
/note="Organ: head; Vector: pPIC1; Site_1: XhoI; Site_
bamHI; Library was kindly generated by Piero Carninci
the RIKEN. The library was normalized and excised usin
re recombinase. Plasmid cDNA library."
61 a 64 c 69 q 70 t
BASE COUNT

```

BASE COUNT  
ORIGIN

|                       |              |                |            |             |
|-----------------------|--------------|----------------|------------|-------------|
| Query Match           | 29.7%        | Score 31.8;    | DB 10;     | Length 264; |
| Best Local Similarity | 67.2%        | Pred. No. 3.8; |            |             |
| Matches 45:           | Conservative | 0;             | Mismatches | 22;         |
|                       |              |                | Models     | 0;          |

Qy 22 tgttgccctggcgcgcaaacggttagagccttcgcaggcacaggttctctaccggattcag 81

Db 101 TGCTGCCAAGGCTGCAAAGGCTGCAGCTGCCGCAATGGAAAGATTGTGGCCGTAA TTGG 160

Qy 82 tgcagtc 88

Db 161 TGCAGTC 167

RESULT 8  
BT239641

|            |  |        |      |        |                 |
|------------|--|--------|------|--------|-----------------|
| LOCUS      | BI239641   | 274 bp | mRNA | linear | EST 12-JUL-2001 |
| DEFINITION | <p>RE36185.Sprime RE Drosophila melanogaster normalized Embryo pFic-1<br/> Drosophila melanogaster cDNA clone RE36185 5 similar to<br/> ATSPsyn-beta: FBan0011154 'enzyme' -AND- 'enzyme' located on: 4<br/> 102F4-102F4; 05/12/2001, mRNA sequence.</p> |        |      |        |                 |

ACCESSION  
VERSION

| KEYWORDS  | SOURCE     | ORGANISM  | REFERENCE          |
|-----------|------------|---|--------------------|
| EST. 1982 | fruit fly. | <i>Drosophila melanogaster</i>  |                    |
|           |            | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. |                    |
|           |            |   | 1 (bases 1 to 274) |

**AUTHORS**

Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C. J., Nunoo, J., Paaleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin, G. M.

TITLE BDGP/HMI RE Drosophila EST Project  
JOURNAL Unpublished (2001)

COMMENT

Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
Plate: RE:361 row: H column: 1  
High quality sequence stop: 252.

FEATURES  
source

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RE36185"
/clone_lib="RE Drosophila melanogaster normalized Embryo

```



```

RESULT 11
BI482355
LOCUS
DEFINITION
  BI482355
  RE65038.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
  Drosophila melanogaster cDNA clone RE65038 5 similar to
  ATPsyn-beta: FBan0011154 GO: [hydrogen-transporting ATP synthase
  (GO:0003936); hydrogen-transporting ATP synthase, catalytic core
  CF(1) (GO:0005754); hydrogen-transporting ATP synthase (GO:0003936
  )] located on: 4 102F4-102F4:: 05/16/2001, mRNA sequence.
ACCESSION
  BI482355
VERSION
  BI482355.1 GI:15324116
KEYWORDS
  EST.
SOURCE
  fruit fly.
ORGANISM
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 287)
REFERENCE
  Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
  , J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
  , R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,
  Mungall, C. J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
  Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin
  , G. M.
  BDGP/HHMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  Plate: RE.650 row: D column: 2
  High quality sequence stop: 221.
FEATURES
  source
    1..287
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone="RE65038"
    /clone_lib="RE Drosophila melanogaster normalized Embryo
    pFlc-1"
    /sex="male and female"
    /dev_stage="0-24 hours mixed stage embryonic"
    /lab_host="DH5-alpha Tona"
    /note="Organ: embryo; Vector: pFlc1; Site:1: XhoI; Site:2:
    BamHI; Library was kindly generated by Piero Carninci at
    the RIKEN. The library was normalized and excised using
    Cre recombinase. Plasmid cDNA library."
  BASE COUNT
    68 a 69 c 79 g 71 t
  ORIGIN
    22 tttgtccctggccgaacgttagaccttcgcacaggcacaggttctgacccgattcag 81
    |||||
    87 TGCTGCCAAGGCTGCRAAGGCTCCAGCTCCGCAATGGAAGATTGTGCCGTAATTGG 146
    |||||
    82 tgcagtc 88
    |||||
    Db 147 TGCAGTC 153

  Query Match 29.7%; Score 31.8; DB 10; Length 287;
  Best Local Similarity 67.2%; Pred. No. 3.9;
  Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

  QY 22 tttgtccctggccgaacgttagaccttcgcacaggcacaggttctgacccgattcag 81
  |||||
  Db 87 TGCTGCCAAGGCTGCRAAGGCTCCAGCTCCGCAATGGAAGATTGTGCCGTAATTGG 146
  |||||
  QY 82 tgcagtc 88
  |||||
  Db 147 TGCAGTC 153

  RESULT 12
  BI242439
  LOCUS
  DEFINITION
    BI242439
    RE39764.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
    Drosophila melanogaster cDNA clone RE39764 5 similar to
    ATPsyn-beta: FBan0011154 'enzyme' -AND- 'enzyme' located on: 4
    102F4-102F4:: 05/12/2001, mRNA sequence.
  ACCESSION
    BI242439
  VERSION
    BI242439.1 GI:14711230
  KEYWORDS
    EST.
  SOURCE
    fruit fly.
  ORGANISM
    Drosophila melanogaster
    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
    Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
    Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
    1 (bases 1 to 306)
  REFERENCE
    Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
    , J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
    , R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,
    Mungall, C. J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
    Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin
    , G. M.
    BDGP/HHMI RE Drosophila EST Project
    Unpublished (2001)
    Contact: Stapleton, M.
    BDGP
    Lawrence Berkeley National Lab
    One Cyclotron Rd, Berkeley, CA 94720, USA
    Fax: 510 486 6798
    Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
    Plate: RE.397 row: F column: 4
    High quality sequence stop: 303.
  FEATURES
    source
      1..306
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone="RE39764"
      /clone_lib="RE Drosophila melanogaster normalized Embryo
      pFlc-1"
      /sex="male and female"
      /dev_stage="0-24 hours mixed stage embryonic"
      /lab_host="DH5-alpha Tona"
      /note="Organ: embryo; Vector: pFlc1; Site:1: XhoI; Site:2:
      BamHI; Library was kindly generated by Piero Carninci at
      the RIKEN. The library was normalized and excised using
      Cre recombinase. Plasmid cDNA library."
    BASE COUNT
      72 a 71 c 85 g 78 t
    ORIGIN
      22 tttgtccctggccgaacgttagaccttcgcacaggcacaggttctgacccgattcag 81
      |||||
      91 TGCTGCCAAGGCTGCRAAGGCTCCAGCTCCGCAATGGAAGATTGTGCCGTAATTGG 150
      |||||
      QY 82 tgcagtc 88
      |||||
      Db 151 TGCAGTC 157

      Query Match 29.7%; Score 31.8; DB 10; Length 306;
      Best Local Similarity 67.2%; Pred. No. 4;
      Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

      QY 22 tttgtccctggccgaacgttagaccttcgcacaggcacaggttctgacccgattcag 81
      |||||
      Db 91 TGCTGCCAAGGCTGCRAAGGCTCCAGCTCCGCAATGGAAGATTGTGCCGTAATTGG 150
      |||||
      QY 82 tgcagtc 88
      |||||
      Db 151 TGCAGTC 157

      RESULT 13
      BI358576
      LOCUS
      DEFINITION
        BI358576
        RE45394.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
        Drosophila melanogaster cDNA clone RE45394 5 similar to
        ATPsyn-beta: FBan0011154 'enzyme' -AND- 'enzyme' located on: 4
        102F4-102F4:: 05/13/2001, mRNA sequence.
      ACCESSION
        BI358576
      VERSION
        BI358576.1 GI:15053030
      KEYWORDS
        EST.
      SOURCE
        fruit fly.
      ORGANISM
        Drosophila melanogaster
        Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
        Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
        Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
        1 (bases 1 to 306)
      REFERENCE
        Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
        , J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George

```



```
/sex="male and female"  
/dev_stage="0 to 24 hours mixed stage embryonic"  
/lab_host="SOLR"  
/note="Organ: embryo; Vector: Bluescript SK; Site_1: EcoRI  
; Site_2: XhoI; Constructed using Stratagene ZAP-CDNA  
Synthesis kit. Oligo dt-primed and directionally cloned at  
EcoRI and XhoI in Bluescript SK(+/-)"  
BASE COUNT      76 a      84 c      89 g      83 t  
ORIGIN
```

```
Query Match      29.7%; Score 31.8; DB 9; Length 332;  
Best Local Similarity 67.2%; Pred. No. 4.1;  
Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
Qy 22 ttgtgacctggccgaacgttagagcttcgcacaggcacaggttctgacccgattcag 81  
Db 109 TGCTGCCAAGGCTGCAAGGCTGCAGCTGCCGCAATGGAAGATTGTGCCCGTAATTGG 168  
Qy 82 tgcagtc 88  
Db 169 TGCACTC 175
```

Search completed: July 15, 2002, 21:48:48  
Job time: 18879 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: July 15, 2002, 22:28:32 ; Search time 2368.24 Seconds  
(without alignments)  
980.832 Million cell updates/sec

Title: US-10-053-641-3  
Perfect score: 111  
Sequence: 1 tctagaggcgcaaaaaatca.....agaacctgcaataatagggc 111

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htgo.inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query  
No. Score Match Length DB ID Description  
-----

|      |      |      |        |    |          |                    |
|------|------|------|--------|----|----------|--------------------|
| 1    | 37   | 33.3 | 195    | 6  | I25751   | 125751 Sequence 27 |
| 2    | 37   | 33.3 | 215    | 6  | A36136   | A36136 Sequence 1  |
| 3    | 37   | 33.3 | 215    | 6  | E08760   | E08760 DNA with Es |
| 4    | 35.8 | 32.3 | 563    | 6  | I25747   | I25747 Sequence 22 |
| 5    | 35.8 | 32.3 | 567    | 6  | A36139   | A36139 Sequence 4  |
| c 6  | 35.2 | 31.7 | 128    | 6  | I41185   | I41185 Sequence 5  |
| 7    | 35.2 | 31.7 | 226    | 12 | SYNHIR   | M26762 Synthetic h |
| 8    | 35.2 | 31.7 | 235    | 6  | I41187   | I41187 Sequence 7  |
| 9    | 35.2 | 31.7 | 235    | 6  | I41188   | I41188 Sequence 8  |
| 10   | 35.2 | 31.7 | 726    | 6  | I41189   | I41189 Sequence 9  |
| c 11 | 35.2 | 31.7 | 726    | 6  | I41190   | I41190 Sequence 10 |
| 12   | 34   | 30.6 | 106    | 6  | I41184   | I41184 Sequence 4  |
| 13   | 33.2 | 29.9 | 152968 | 2  | AC073669 | AC073669 Mus muscu |
| 14   | 33   | 29.7 | 59     | 6  | I63542   | I63542 Sequence 5  |
| 15   | 33   | 29.7 | 59     | 6  | I63546   | I63546 Sequence 9  |
| c 16 | 32.2 | 29.0 | 207374 | 2  | AC105822 | AC105822 Rattus no |
| 17   | 32   | 28.8 | 279    | 6  | A01139   | A01139 Fuson DNA   |
| 18   | 32   | 28.8 | 279    | 6  | I26634   | I26634 Sequence 3  |
| c 19 | 31.6 | 28.5 | 146809 | 2  | AC091872 | AC091872 Homo sapi |
| c 20 | 31.6 | 28.5 | 155150 | 9  | AC027332 | AC027332 Homo sapi |
| 21   | 31.4 | 28.3 | 59     | 6  | I63548   | I63548 Sequence 11 |
| 22   | 31.2 | 28.1 | 250    | 14 | AF067248 | AF067248 Tomato mo |
| 23   | 31   | 27.9 | 195    | 6  | A43186   | A43186 Sequence 7  |
| 24   | 31   | 27.9 | 195    | 6  | I62835   | I62835 Sequence 7  |
| 25   | 31   | 27.9 | 198    | 12 | AF274588 | AF274588 Synthetic |
| 26   | 31   | 27.9 | 273    | 12 | AF284216 | AF284216 Synthetic |
| 27   | 31   | 27.9 | 2157   | 8  | AB000799 | AB000799 Arabidops |
| 28   | 31   | 27.9 | 2240   | 8  | AY069917 | AY069917 Arabidops |
| c 29 | 31   | 27.9 | 23404  | 2  | AC017160 | AC017160 Drosophil |
| c 30 | 31   | 27.9 | 24745  | 8  | AC068073 | AC068073 Arabidops |
| c 31 | 31   | 27.9 | 192366 | 3  | AC008369 | AC008369 Drosophil |
| c 32 | 31   | 27.9 | 258873 | 3  | AE003810 | AE003810 Drosophil |
| c 33 | 30.8 | 27.7 | 178    | 6  | A13978   | A13978 Synthetic D |
| 34   | 30.8 | 27.7 | 180    | 6  | A02448   | A02448 Nucleotide  |
| c 35 | 30.8 | 27.7 | 180    | 6  | A02449   | A02449 Nucleotide  |
| 36   | 30.8 | 27.7 | 180    | 6  | A13977   | A13977 Synthetic D |
| 37   | 30.8 | 27.7 | 212    | 6  | A13382   | A13382 DNA sequenc |
| c 38 | 30.8 | 27.7 | 212    | 6  | A13383   | A13383 DNA sequenc |
| 39   | 30.8 | 27.7 | 212    | 6  | A34619   | A34619 Synthetic h |
| c 40 | 30.8 | 27.7 | 212    | 6  | A34620   | A34620 Synthetic h |
| 41   | 30.8 | 27.7 | 212    | 6  | E00711   | E00711 Synthetic D |
| 42   | 30.8 | 27.7 | 225    | 6  | A04613   | A04613 Nucleotide  |
| c 43 | 30.8 | 27.7 | 225    | 6  | A04614   | A04614 Nucleotide  |
| 44   | 30.8 | 27.7 | 226    | 6  | A04615   | A04615 Nucleotide  |
| 45   | 30.8 | 27.7 | 238    | 6  | A04616   | A04616 HV1 gene. 8 |

ALIGNMENTS

|          |            |  |                                     |      |        |                 |
|----------|------------|--|-------------------------------------|------|--------|-----------------|
| RESULT 1 | I25751     | Sequence 27 from patent US 5552299.  | 195 bp                              | DNA  | linear | PAT 07-OCT-1996 |
| I25751   | LOCUS      | I25751   | Sequence 27 from patent US 5552299. |      |        |                 |
|          | DEFINITION | I25751   |                                     |      |        |                 |
|          | ACCESSION  | I25751   |                                     |      |        |                 |
|          | VERSION    | I25751.1   | GI:1605621                          |      |        |                 |
|          | KEYWORDS   | Unknown.   |                                     |      |        |                 |
|          | SOURCE     | Unknown.   |                                     |      |        |                 |
|          | ORGANISM   | Unclassified.  |                                     |      |        |                 |
|          | REFERENCE  | 1 (bases 1 to 195)   |                                     |      |        |                 |
|          | AUTHORS    | Ott, Ian., Klupp, T., Molnar, Ian., Patthy, Aas., Barta, Ian., Bark, O n<br>ee T oth, Z., Ambrus, Gabor., Sal at, Janos., Tegdes, Ao.,<br>Moravcsik, I., Esg ud, C., Albrecht, Karnly., K oncz Ol, Kalman.,<br>Vincze, A., Barab as, E., M at e, Gorgy., Kiss, Gorgy. B., Kiss, Peter.,<br>P Oly, Kalman., Erdei, Janos., Guly as, E. and Zilahi, E.<br>Plasmids and process for producing recombinant desulphatohirudin<br>HV-1 peptides<br>Patent: US 5552299-A 27 03-SEP-1996;<br>Location/Qualifiers<br>1. .195<br>/organism="unknown" |                                     |      |        |                 |
|          | TITLE      |  |                                     |      |        |                 |
|          | JOURNAL    |  |                                     |      |        |                 |
|          | FEATURES   |  |                                     |      |        |                 |
|          | source     |  |                                     |      |        |                 |
|          | BASE COUNT | 58 a   | 40 c                                | 45 g | 51 t   | 1 others        |

1





```

FEATURES                               Location/Qualifiers
    source                            1..726
        /organism="unknown"
BASE COUNT      149 a       126 c       171 g       280 t
ORIGIN

Query Match          31.7%; Score 35.2; DB 6; Length 726;
Best Local Similarity 83.3%; Pred. No. 0.57;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY   1  tctagaggcgaataaatcaatcgcttactggcgaaagtactccacaa 48
     |||  ||||| ||| ||||| ||||| ||||| ||||| |||
Db   108 TCTGACGGCGAAAAAACCAAGTCGTACTGGCGAAGGTACCCCGAAA 61

RESULT 12
LOCUS      I41184
DEFINITION Sequence 4 from patent US 5624822.
ACCESSION  I41184
VERSION     I41184.1 GI:2081774
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 106)
AUTHORS    Koerwer,W.
TITLE      Hirudin fusion proteins and preparation of hirudin
JOURNAL    Patent: US 5624822-A 4 29-APR-1997;
FEATURES   Location/Qualifiers
            1..106
            /organism="unknown"
BASE COUNT 36 a       28 c       28 g       14 t
ORIGIN

Query Match          30.6%; Score 34; DB 6; Length 106;
Best Local Similarity 88.1%; Pred. No. 1.5;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY   7  gccgaaaaaatcaatgcgttactggcgaaagtactccacaa 48
     ||||| ||||| ||| ||||| ||||| ||||| ||||| |||
Db   4  GCGGAAAAAACCAAGTCGTACTGGCGAAGGTACCCCGAAA 45

RESULT 13
LOCUS      AC073669
DEFINITION Mus musculus clone CT7-273D11, WORKING DRAFT SEQUENCE, 39 unordered pieces.
ACCESSION  AC073669
VERSION    AC073669.1 GI:8810286
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      Sequencing of Mouse
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 152968)
AUTHORS    DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL    Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT    -----Genome Center
           Center: Joint Genome Institute
           Center Code: JGI
           Web site: http://www.jgi.doe.gov
           -----
           Project Information

```

Center Project Name: 1058096  
Center clone name: RG-MBAC\_273D11

# Summary Statistics

Consensus quality: 118135 bases at least Q40  
Consensus quality: 134236 bases at least Q30  
Consensus quality: 137982 bases at least Q20  
Estimated insert size: 100000; pulse field gel estimation  
Estimated insert size: 149168; sum-of-contigs estimation  
Quality coverage: 7.41 in Q20 bases; pulse field gel estimation  
Quality coverage: 4.96 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 39 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1093: contig of 1093 bp in length  
\* 1094 1193: gap of unknown length  
\* 1194 2195: contig of 1002 bp in length  
\* 2196 2295: gap of unknown length  
\* 2296 3335: contig of 1040 bp in length  
\* 3336 3436: gap of unknown length  
\* 3437 4873: contig of 1438 bp in length  
\* 4874 5084: contig of 1111 bp in length  
\* 5085 6185: gap of unknown length  
\* 6186 7743: contig of 1558 bp in length  
\* 7744 8945: contig of 1103 bp in length  
\* 8946 10221: contig of 1075 bp in length  
\* 10222 11476: contig of 1255 bp in length  
\* 11477 12700: contig of 1124 bp in length  
\* 12701 12988: contig of 1188 bp in length  
\* 12989 14088: contig of 1095 bp in length  
\* 14089 15183: gap of unknown length  
\* 15184 16468: contig of 1185 bp in length  
\* 16469 18335: gap of unknown length  
\* 18336 20772: contig of 2237 bp in length  
\* 20773 22314: contig of 1341 bp in length  
\* 22315 23495: contig of 1182 bp in length  
\* 23496 25217: contig of 1622 bp in length  
\* 25218 27409: contig of 2092 bp in length  
\* 27410 28855: contig of 1356 bp in length  
\* 28856 31579: contig of 2614 bp in length  
\* 31580 34419: contig of 2740 bp in length  
\* 34420 38002: contig of 3483 bp in length  
\* 38003 42293: contig of 4191 bp in length  
\* 42294 47356: contig of 4963 bp in length  
\* 47357 51699: contig of 4243 bp in length  
\* 51700 55044: contig of 3245 bp in length  
\* 55045 55144: gap of unknown length

\* 55145 59713: contig of 4569 bp in length  
\* 59714 64897: gap of unknown length  
\* 64898 69683: contig of 5084 bp in length  
\* 69684 74689: gap of unknown length  
\* 74690 80906: contig of 4686 bp in length  
\* 80907 91022: gap of unknown length  
\* 91023 100746: contig of 4906 bp in length  
\* 100747 107914: gap of unknown length  
\* 107915 118004: contig of 7069 bp in length  
\* 118005 131953: gap of unknown length  
\* 131954 140157: contig of 13948 bp in length  
\* 140158 140257: gap of unknown length  
\* 140258 152968: contig of 8105 bp in length  
\* 152969 172711: contig of 12711 bp in length.

## FEATURES

source  
1. .152968  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="C7-273D11"  
/clone\_lib="C1BCJ7 mouse BAC library"  
BASE COUNT 39470 a 32372 c 31940 g 45341 t 3845 others  
ORIGIN

Query Match 29.9%; Score 33.2; DB 2; Length 152968;  
Best Local Similarity 59.6%; Pred. No. 2.3; Indels 0; Gaps 0;  
Matches 56; Conservative 0; Mismatches 38;  
Qy 4 agagcgcaaaaaaataatcgcttactgcgaagtgactccacaggtttgaatcccg 63  
DB 38124 ATACACTAAAGACACATCCAGCCAGGCTACTATACCTGCCAAACTCTCAATTACCA 38183  
Qy 64 aagaatagaacccgagctctccgcagacagaacc 97  
DB 38184 TAGAGGAGAACCAAGTATTCCACGACAAACC 38217

RESULT 14  
LOCUS I63542 I63542 59 bp DNA linear PAT 07-OCT-1997  
DEFINITION Sequence 5 from patent US 5663141.  
ACCESSION I63542  
VERSION I63542.1 GI:2481115  
KEYWORDS  
SOURCE Unknown.  
ORGANISM  
Unclassified.  
REFERENCE 1 (bases 1 to 59)  
AUTHORS Kurfuerst, M., Ruebmaier, K., Schmied, B., Koerwer, W., Schweden, J. and  
Hoeffken, H. Wolfgang.  
TITLE Hirusin/polyalkylene glycol conjugates and hirusin mutins  
JOURNAL Patent: US 5663141-A 5 02-SEP-1997;  
FEATURES  
Location/Qualifiers  
source  
1. .59  
/organism="unknown"  
BASE COUNT 19 a 14 c 16 g 10 t  
ORIGIN

Query Match 29.7%; Score 33; DB 6; Length 59;  
Best Local Similarity 87.8%; Pred. No. 3.2; Indels 0; Gaps 0;  
Matches 36; Conservative 0; Mismatches 5;  
Qy 1 tctagagcgcaaaaaaataatcgcttactgcgaaggtac 41

Db 19 TCTGACGGGCAAAAACCAAGTCGTTACTGCGAAGGTAC 59  
|||||  
RESULT 15  
I63546  
LOCUS I63546 59 bp DNA linear PAT 07-OCT-1997  
DEFINITION Sequence 9 from patent US 5663141.  
ACCESSION I63546  
VERSION I63546.1 GI:2481119  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 59)  
AUTHORS Kurfuerst, M., Ruebsamen, K., Schmied, B., Koerwer, W., Schweden, J. and  
Hoeffken, H. Wolfgang.  
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins  
JOURNAL Patent: US 5663141-A 9 02-SEP-1997;  
FEATURES Location/Qualifiers  
source 1..59  
BASE COUNT 20 a 14 c 15 g 10 t  
ORIGIN  
Query Match 29.7%; Score 33; DB 6; Length 59;  
Best Local Similarity 87.8%; Pred. No. 3.2;  
Matches 36; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 tctagaggcgaaaaaatcaatgcgttactgcggaaggtac 41  
|||||  
Db 19 TCTAAAGGCCAGAAAAACCAAGTCGTTACTGCGAAGGTAC 59

Search completed: July 15, 2002, 22:28:44  
Job time: 18970 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:47 ; Search time 758.37 Seconds  
(without alignments)  
251.299 Million cell updates/sec

Title: US-10-053-641-3

Perfect score: 111

Sequence: 1 tctagaggcgaaataca.....agaacctgcaataataggc 111

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

```

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | ID   | Description |
|------------|-------|-------------|------|-------------|
| 1          | 37    | 33.3        | 219  | 15 AAQ54995 |
| 2          | 35.8  | 32.3        | 567  | 15 AAQ54997 |
| 3          | 35.2  | 31.7        | 726  | 12 AAQ12379 |
| 4          | 33.2  | 29.9        | 468  | 16 AAQ93244 |
| 5          | 32    | 28.8        | 279  | 12 AAQ13831 |
| 6          | 31    | 27.9        | 195  | 16 AAQ81528 |
| 7          | 31    | 27.9        | 1984 | 21 AAZ98321 |
| 8          | 31    | 27.9        | 2157 | 21 AAZ51103 |
| 9          | 30.8  | 27.7        | 227  | 22 AAF61507 |

|    |      |      |       |             |                    |
|----|------|------|-------|-------------|--------------------|
| 10 | 30.8 | 27.7 | 229   | 9 AAN81294  | Sequence encoding  |
| 11 | 30.8 | 27.7 | 238   | 9 AAN81295  | Sequence encoding  |
| 12 | 30.6 | 27.6 | 231   | 7 AAN60354  | Desulphatohirudin  |
| 13 | 30.4 | 27.4 | 195   | 13 AAQ27601 | Encodes hirudin de |
| 14 | 30.4 | 27.4 | 196   | 8 AAN70319  | Sequence encoding  |
| 15 | 30.4 | 27.4 | 208   | 10 AAN91851 | Synthetic DNA enco |
| 16 | 30.4 | 27.4 | 217   | 7 AAN60355  | Desulphatohirudin  |
| 17 | 30.4 | 27.4 | 217   | 8 AAN70323  | Sequence of the de |
| 18 | 30.4 | 27.4 | 224   | 7 AAN60746  | DNA encoding hirud |
| 19 | 30.4 | 27.4 | 276   | 13 AAQ27600 | Encodes hirudin de |
| 20 | 30.4 | 27.4 | 525   | 11 AAQ06873 | Secretion plasmid  |
| 21 | 30.4 | 27.4 | 526   | 11 AAQ03237 | Sequence of DNA in |
| 22 | 30   | 27.0 | 358   | 6 AAN50397  | Synthetic hirudin  |
| 23 | 29.8 | 26.8 | 43    | 14 AAQ43368 | Hirudin oligonucle |
| 24 | 29.4 | 26.5 | 257   | 10 AAN91836 | Synthetic HVL gene |
| 25 | 29.4 | 26.5 | 257   | 10 AAN91867 | Synthetic hirudin  |
| 26 | 29.4 | 26.5 | 1082  | 15 AAQ64146 | CUP1 promoter, PHO |
| 27 | 29.4 | 26.5 | 1082  | 15 AAQ45312 | Yeast CUP1 promote |
| 28 | 29.4 | 26.5 | 1130  | 16 AAQ81527 | PJDB207/GAPFL-YHIR |
| 29 | 29.2 | 26.3 | 24158 | 21 AAN81532 | N. meningitidis pa |
| 30 | 29.2 | 26.3 | 34980 | 21 AAF21609 | Neisseria meningit |
| 31 | 28.8 | 25.9 | 198   | 13 AAQ24986 | Sequence encoding  |
| 32 | 28.8 | 25.9 | 198   | 13 AAQ63876 | Hirudin variant.   |
| 33 | 28.8 | 25.9 | 210   | 12 AAQ10452 | HV-1 gene. Synthe  |
| 34 | 28.2 | 25.4 | 201   | 12 AAQ12153 | Synthetic hirudin  |
| 35 | 28.2 | 25.4 | 201   | 13 AAQ25184 | Hirudin HV-1. Syn  |
| 36 | 28.2 | 25.4 | 420   | 12 AAQ12155 | Factor Xa-cleavabl |
| 37 | 28.2 | 25.4 | 1458  | 12 AAQ12162 | Factor Xa-cleavabl |
| 38 | 28.2 | 25.4 | 1467  | 12 AAQ12490 | Factor Xa-cleavabl |
| 39 | 28   | 25.2 | 183   | 12 AAQ14927 | Partial Hirudin HV |
| 40 | 28   | 25.2 | 7157  | 22 ABA19506 | Human nervous syst |
| 41 | 27.8 | 25.0 | 944   | 21 ABA1426  | Arabidopsis thalia |
| 42 | 27.8 | 25.0 | 1335  | 21 AAC36967 | Arabidopsis thalia |
| 43 | 27.4 | 24.7 | 422   | 21 AAC38173 | Arabidopsis thalia |
| 44 | 27.4 | 24.7 | 959   | 21 AAC33229 | Arabidopsis thalia |
| 45 | 27.4 | 24.7 | 8900  | 22 AAL36092 | Human musculoskele |

#### ALIGNMENTS

RESULT 1

AAQ54995  
ID AAQ54995 standard; DNA; 219 BP.

XX AAQ54995;

AC AAQ54995;

DT 11-JUL-1994 (first entry)

XX Desulphatohirudin gene with E. coli-Saccharomyces codon usage.

XX Hirudin; HV-1; E. coli; codon usage: desulphatohirudin; HV01; 33asp;

XX 33 Asp; biological activity; thrombosis; thromboembolism; ds.

XX Hirudo medicinalis.

XX Key Location/Qualifiers

XX CDS 14..208

XX /\*tag= a

XX /product= Hirudin\_HV-1

XX EP576792-A.

XX 05-JAN-1994.

XX 13-APR-1993; 93EP-0105848.

XX 09-APR-1992; 92HU-0001200.

XX (BIOG ) BIOGAL GYOGYSZERVAR.

XX Albrecht K, Ambrus G, Barabas E, Barta I, Berk T;

XX Botond K, Egyued C, Erdei J, Kiss P, Klupp T, Koenczoel K;





```

PR 22-MAR-1990; 90GB-0006400.
XX
PA (CIBA ) CIBA GEIGY AG.
XX
XX Surli B, Schmitz A;
PI
XX WPI; 1991-290162/40.
DR
DR P-PSDB; AARL14151.
XX
XX Hybrid vectors for expression of polypeptide(s) - comprise DNA
PT from plasmids obtd. from Lactococcus lactis, esp. major secretion
PT prod. gene fragments.
XX
XX Disclosure: Page 26; 36pp; English.
XX
XX The sequence (SEQ ID NO:2) was obtd. by ligation of the signal
CC peptide encoding DNA from L. lactis (from pUCRS, DSM 5803) and the
CC coding region for desulphatohirudin (from pML310, EP-168342).
CC The presence of the signal peptide results in the secretion of
CC expressed hirudin into the supernatant.
CC See also AAQ13830.
XX
XX Sequence 279 BP; 76 A; 66 C; 68 G; 69 T; 0 other;
SQ
Query Match 28.8% Score 32; DB 12; Length 279;
Best Local Similarity 79.2%; Pred.No. 0.12; Indels 0; Gaps 0;
Matches 38; Conservative 0; Mismatches 10;
QY 1 tctagcgcgaataaatcaatgcttactggcgaaagtactccacaa 48
Db 175 tctgacggtgaaaaaacaccagtgcgttacgccgaaggtaacctcgaaa 222
RESULT 6
AAQ81528
ID ID AAQ81528 standard; DNA; 195 BP.
XX AC AAQ81528;
XX
DT 29-SEP-1995 (first entry)
XX
XX Leech hirudin mutcin 33-Cys HV1.
DE
XX Hirudin mutant; leech; anticoagulant; antithrombitic; Cys 33 HV1; ss.
KW
XX Synthetic.
OS
XX Key Location/Qualifiers
FH FT CDS 1..195
FT FT /*tag= a
XX
PN WO9504823-A.
XX
XX 16-FEB-1995.
XX
XX 23-JUL-1994; 94WO-EF02438.
XX
XX 04-AUG-1993; 93EP-0810552.
XX
XX (CIBA ) CIBA GEIGY AG.
PA (UPCG-) UPC GEN-PHARMA AG.
XX
XX Chang JY, Grossenbacher H, Marki W, Maerki W;
PI
XX WPI; 1995-090901/12.
XX
XX P-PSDB; AAR69100.
XX
XX Conjugate consisting of 2-4 residues of desulphatohirudin muteins
PT - having hirudin activity, used in compns. to treat thromboses
PT and embolism
XX
XX Claim 34; Page 38; 44pp; English.
PS

```







|  |   |
|--|---|
| XX   | Liersch M, Rink H, Markl W, Grutter MG, Meyhack B;                                      |
| PI   | WPI; 1986-015589/03.  |
| DR   | P-PSDB; AAP60395.   |
| XX   |   |
| PT   | DNA sequences coding for hirudin and derivs. - and expression                           |
| PT   | vectors, transformed cells, monoclonal antibodies and hybridomas,                       |
| PT   | useful as thrombin inhibitor.   |
| XX   |   |
| PPS  | Disclosure; Page 8; 123pp; German.  |
| XX   |   |
| CC   | The sequence pref. comprises an EcoRI restriction enzyme site at the                    |
| CC   | 5'-end and a BamHI restriction enzyme site at the 3'-end. See                           |
| CC   | AAN60355 for the preferred sequence. The sequence may be introduced                     |
| CC   | into a vector for the transformation of hosts, e.g. E. coli. The                        |
| CC   | hirudin can thus be prepared on a large scale. The product and its                      |
| CC   | derivs. are thrombin inhibitors, useful in antitoxicant therapy,                        |
| CC   | esp. when injected at doses of 0.01-0.05 mg/kg.   |
| XX   |   |
| SQ   | Sequence 231 BP; 43 A; 16 C; 41 G; 22 T; 109 other;                                     |
|  |   |
| Query Match            27.6%; Score 30.6; DB 7; Length 231;      |   |
| Best Local Similarity   64.3%; Pred. No. 0.35;                   |   |
| Matches   27; Conservative   7; Mismatches   8; Indels   0; Gaps |   |
| QY   | 7 ggcgaaaaaatcaatgcgttacttcggcgaaaggctaccacaa 48<br>   :::  :::  :::  :::  :::  :::  :: |
| Db   | 118 ggngamaamaaycmtgytgtnacggngangnacnccnaam 159  |
|  |   |
| RESULT 13  |   |
| AAQ27601   |   |
| ID   | AAQ27601 standard; DNA: 195 BP.   |
| AC   | AAQ27601;   |
| XX   |   |
| DT   | 04-FEB-1993 (first entry)   |
| XX   |   |
| DE   | Encodes hirudin derived thombin inhibitor #2.   |
| XX   |   |
| KW   | bleeding time; clotting time; anti-thrombotic;  |
| KW   | congestive phlebothrombosis.  |
| OS   | Hirudo medicinalis.   |
| PN   | JP04197184-A.   |
| PD   | 16-JUL-1992.  |
| XX   |   |
| PF   | 28-NOV-1990; 90JP-0323133.  |
| XX   |   |
| PR   | 28-NOV-1990; 90JP-0323133.  |
| XX   |   |
| PA   | (MITK ) MITSUI TOATSU CHEM INC.   |
| XX   |   |
| DR   | WPI; 1992-288441/35.  |
| XX   |   |
| PT   | P-PSDB; AAR26314.   |
| XX   |   |
| PT   | Thrombin-inhibiting polypeptide(s) more potent than hirudin HVI -                       |
| PT   | produced by transformed Bacillus Subtilis, for treating                                 |
| XX   | thrombosis  |
| XX   |   |
| PS   | Claim 6; Page 1; 24pp; Japanese.  |
| XX   |   |
| XX   | This sequence encodes a novel, hirudin derived thrombin inhibitor.                      |
| XX   |   |
| SQ   | Sequence 195 BP; 76 A; 24 C; 42 G; 53 T; 0 other;                                       |
|  |   |
| Query Match            27.4%; Score 30.4; DB 13; Length 195;     |   |
| Best Local Similarity   77.1%; Pred. No. 0.39;                   |   |

```

XX Synthetic DNA encoding desulphatohirudin.
DE Desulphatohirudin; synthetic gene; thrombin inhibitor; HV-1 gene;
KW plasmid p3010; plasmid p4014.
XX
XX JP01247092-A.
XX
XX 02-OCT-1989.
XX
XX 29-MAR-1988; 88JP-0073200.
XX
XX 29-MAR-1988; 88JP-0073200.
XX
XX (MITK ) MITSUI TOATSU CHEM. INC.
XX
XX WPI; 1989-330037/45.
XX
XX Prepn. of desulphatohirudin protein used as thrombin inhibitor - by
PT introducing specific DNA into expression vector, inserting obtd.
PT recombinant plasmid into host and incubating.
XX
XX Claim 1; page 485; 6pp; Japanese.
XX
XX The synthec gene is constructed by dividing the hirudin HV-1 gene into
CC 7 parts and ligating to form the gene. It is inserted into pBR322 to
CC give plasmid p3010. This is inserted into expression vector pKK223-3 to
CC give plasmid p4014. This is used to transform E.coli. It produces
CC desulphatohirudin, which is active as a thrombin inhibitor. It allows
CC easy mass prodn.
XX
XX Sequence 208 BP; 82 A; 23 C; 43 G; 60 T; 0 other;
SQ

```

Query Match 27.4%; Score 30.4; DB 10; Length 208;  
 Best Local Similarity 77.1%; Pred. No. 0.4;  
 Matches 37; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```

Qy 1 tctagaggcgagaaataatcggttactgcggaagggtactccacaa 48
   ||| || ||||| ||||| ||||| || ||||| || ||
Db 100 tctgatggagagaaataatcggttactagaggagagaaacccgaaa 147

```

Search completed: July 15, 2002, 22:44:48  
 Job time: 10061 sec













RESULT 14  
5180668-10  
; Patent No. 5180668  
; APPLICANT: CRAUSE, PETER; HABERMANN, PAUL; TRIPIER, DOMINIQUE  
; TITLE OF INVENTION: HIRUDIN DERIVATIVE  
; NUMBER OF SEQUENCES: 10  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/295,422  
; FILING DATE: 10-JAN-1989  
; SEQ ID NO:10:  
; LENGTH: 212  
5180668-10

| Query Match           | 27.7%           | Score 30.8;                                 | DB 2;  | Length 227; |
|-----------------------|-----------------|---|--------|-------------|
| Best Local Similarity | 83.3%           | Pred. No. 0.03;                             |        |             |
| Matches 35;           | Conservative 0; | Mismatches 7;                               | Indels |             |
| Qy                    | 7               | ggcgaaaaaatcaatcgctgttactggcgaggttactccacaa | 48     |             |
| Db                    | 97              | GGTGAAGAAGCAACAGTCGCTTACTGGCGAAGGTACCCGAAA  | 138    |             |

Search completed: July 15, 2002, 22:31:48  
Job time: 18474 sec





GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:48:48 ; Search time 6165.88 Seconds

(without alignments)  
242.976 Million cell updates/sec

Title: US-10-053-641-3

Perfect score: 111

Sequence: 1 tctagaggcgaaaaataca.....agaacctgcaataatagggc 111

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estnu:\*
- 5: em\_estcov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID        | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 30.8  | 27.7        | 455    | AQ704885  | AQ704885 HS_5521_B |
| 2          | 30.2  | 27.2        | 346    | D48127    | D48127 RICS14185A  |
| 3          | 30    | 27.0        | 849    | CNS08YL0  | AL421162 T7 end of |
| 4          | 29.6  | 26.7        | 369    | D47621    | D47621 RICS13224A  |
| 5          | 29    | 26.1        | 167    | BH005024  | BH005024 BMBAC08P1 |
| 6          | 29    | 26.1        | 241    | BG907200  | BG907200 Talr1157H |
| 7          | 29    | 26.1        | 287    | BI233876  | BI233876 949037C04 |
| 8          | 29    | 26.1        | 409    | AJ234149  | AJ234149 B927 133  |
| 9          | 29    | 26.1        | 421    | T76275    | T76275 11053 Lambd |
| 10         | 29    | 26.1        | 442    | BI478547  | BI478547 949070G02 |
| 11         | 29    | 26.1        | 551    | BG316445  | BG316445 947023E01 |
| 12         | 29    | 26.1        | 566    | BI233858  | BI233858 949037A03 |
| 13         | 28.8  | 25.9        | 357    | TA62H05P  | AL465105 T. brucei |
| 14         | 28.8  | 25.9        | 553    | TA278H10Q | AL486158 T. brucei |
| 15         | 28.8  | 25.9        | 573    | TA174F12Q | AL474540 T. brucei |
| 16         | 28.8  | 25.9        | 660    | AL554638  | AL554638 AL554638  |
| 17         | 28.6  | 25.8        | 328    | D47040    | D47040 RICS12113A  |

|      |      |      |     |    |          |
|------|------|------|-----|----|----------|
| C 18 | 28.6 | 25.8 | 343 | 10 | D46970   |
| C 19 | 28.4 | 25.6 | 463 | 10 | T21551   |
| C 20 | 28.4 | 25.6 | 523 | 9  | AA855282 |
| C 21 | 28.4 | 25.6 | 655 | 10 | BI946448 |
| C 22 | 28.2 | 25.4 | 287 | 10 | D47557   |
| C 23 | 28.2 | 25.4 | 499 | 10 | BE411902 |
| C 24 | 28.2 | 25.4 | 502 | 10 | R90001   |
| C 25 | 28.2 | 25.4 | 927 | 10 | BI106318 |
| C 26 | 28   | 25.2 | 405 | 10 | D48111   |
| C 27 | 28   | 25.2 | 453 | 10 | T44933   |
| C 28 | 28   | 25.2 | 480 | 9  | AA873427 |
| C 29 | 28   | 25.2 | 507 | 12 | BH504679 |
| C 30 | 28   | 25.2 | 525 | 10 | N38188   |
| C 31 | 28   | 25.2 | 529 | 10 | N65210   |
| C 32 | 28   | 25.2 | 537 | 10 | H37346   |
| C 33 | 28   | 25.2 | 545 | 10 | H77106   |
| C 34 | 28   | 25.2 | 548 | 10 | N37528   |
| C 35 | 28   | 25.2 | 570 | 10 | BF317079 |
| C 36 | 28   | 25.2 | 603 | 10 | BE468512 |
| C 37 | 28   | 25.2 | 651 | 10 | BM359653 |
| C 38 | 28   | 25.2 | 915 | 12 | CNS034R1 |
| C 39 | 27.8 | 25.0 | 334 | 9  | AV527342 |
| C 40 | 27.8 | 25.0 | 342 | 9  | AV528031 |
| C 41 | 27.8 | 25.0 | 356 | 9  | BE230770 |
| C 42 | 27.8 | 25.0 | 377 | 9  | AV526530 |
| C 43 | 27.8 | 25.0 | 422 | 12 | AQ631826 |
| C 44 | 27.8 | 25.0 | 463 | 10 | T04150   |
| C 45 | 27.8 | 25.0 | 465 | 9  | AV526935 |

## ALIGNMENTS

## RESULT 1

AQ704885 455 bp DNA linear GSS 07-JUL-1999  
 HS\_5521\_B2\_B09\_T7A RPCI-11 Human Male BAC Library Homo sapiens  
 genomic clone Plate-1097 Col-18 Row=D, DNA sequence.

AQ704885  
 GSS.  
 human.

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
 1 (bases 1 to 455)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.

TITLE  
 Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome

JOURNAL  
 PROC. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE  
 Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallaceu.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pletier de Jong  
 (pletier@jeong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
 or from Resear h Genetics (inforesgen.com). BAC end Web Server:  
 http://www.htsc.washington.edu  
 Plate: 1097 row: D column: 18  
 Seg primer: T7  
 Class: BAC ends  
 High quality sequence stop: 455.  
 Location/Qualifiers  
 1. 455  
 /organism="Homo sapiens"

## FEATURES

source





[illegible]



|                       |  |  |                                  |
|-----------------------|--|--|----------------------------------|
| KEYWORDS              | EST.   | thale cress.   |                                  |
| SOURCE                | ORGANISM   | Arabidopsis thaliana   |                                  |
| REFERENCE             |  | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  |                                  |
| AUTHORS               |  | 1 (bases 1 to 421)<br>Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E., and Somerville, C.   |                                  |
| TITLE                 |  | Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones  |                                  |
| JOURNAL               |  | Plant Physiol. 106, 1241-1255 (1994)   |                                  |
| MEDLINE               |  | 95148729   |                                  |
| COMMENT               |  | On Apr 14, 1993 this sequence version replaced gi:693037.  |                                  |
| FEATURES              | source   | <p>Contact: Thomas Newman<br/>MSU-DOE Plant Research Laboratory<br/>Michigan State University<br/>MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI<br/>Tel: 517-353-0854<br/>Fax: 517-353-9168<br/>Email: 22313tcn@bm.cl.msu.edu<br/>Seq primer: T7 dye primer.<br/>Location/Qualifiers<br/>1..421<br/>/organism="Arabidopsis thaliana"<br/>/strain="var columbia"<br/>/db_xref="taxon:3702"<br/>/clone="148P107"<br/>/clone_lib="Lambda-PRL2"<br/>/note="Vector: lambda zip-Lox; Site_1: Sal; Site_2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques. The vector is BRL's lambda zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."</p> |                                  |
| BASE COUNT            | 106 a  | 101 c  | 102 g                            |
| ORIGIN                |  | 96 t   | 16 others                        |
| Query Match           | 26.1%;   | Score 29;  | DB 10; Length 421;               |
| Best Local Similarity | 62.0%;   | Pred. No. 44;  |                                  |
| Matches               | 44;  | Conservative 0;  | Mismatches 27; Indels 0; Gaps 0; |
| QY                    | 34   | gaaggtactccacaagcttgaatcccggaagaataagaaacccgagtcctccgacgacag   | 93                               |
| Db                    | 284  | GGAGATACTCCCGCGTTTGCTTAACCAACCAAGGATCGAACCCGTAGTCCCGACAAATG  | 225                              |
| QY                    | 94   | aacctgcaata 104  |                                  |
| Db                    | 224  | AACCGTCGANA 214  |                                  |
| RESULT 10             |  |  |                                  |
| LOCUS                 | BI478547   | 442 bp   | mrna linear EST 28-AUG-2001      |
| DEFINITION            | 949070G02.y1 949 - Juvenile leaf and shoot cDNA from Steve Moose   |  |                                  |
| ACCESSION             | BI478547   | zea mays cDNA, mRNA sequence.  |                                  |
| VERSION               | BI478547.1   | GI:15323491  |                                  |
| KEYWORDS              | EST.   |  |                                  |
| SOURCE                | zea mays.  |  |                                  |
| ORGANISM              | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC |  |                                  |
| REFERENCE             | 1 (bases 1 to 442)   |  |                                  |
| AUTHORS               | Walbot, V.   |  |                                  |

|                       |  |  |                                  |
|-----------------------|--|--|----------------------------------|
| TITLE                 | Maize ESTs from various cDNA libraries sequenced at Stanford University  |  |                                  |
| JOURNAL               | Unpublished (1999)   |  |                                  |
| COMMENT               | Contact: Walbot V<br>Department of Biological Sciences<br>Stanford University<br>855 California Ave, Palo Alto, CA 94304, USA<br>Tel: 650 723 2227<br>Fax: 650 725 8221<br>Email: walbot@stanford.edu<br>Plate: 949070 row: G column: 02.<br>Location/Qualifiers<br>1..442<br>/organism="Zea mays"<br>/cultivar="W64A"<br>/db_xref="taxon:4577"<br>/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve Moose"<br>/tissue_type="Immature leaf primordium and vegetative meristem"<br>/dev_stage="4 stages from 3-13 days after imbibing"<br>/lab_host="E. coli XL0LR"<br>/note="Organ: Juvenile vegetative shoots; Vector: pAD-GAL4-2.1; Site_1: ECORI; Site_2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcorI (5') and XhoI (3') directional cloning into Lambda Hybrizap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing - includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developing leaves 4-5 and the vegetative apex. Including adult leaf primordia, 9 days after sowing. 4. Partially expanded and greening leaves 4-5 at 13 days after sowing." |  |                                  |
| BASE COUNT            | 67 a   | 164 c  | 144 g                            |
| ORIGIN                |  | 67 t   |                                  |
| Query Match           | 26.1%;   | Score 29;  | DB 10; Length 442;               |
| Best Local Similarity | 63.8%;   | Pred. No. 45;  |                                  |
| Matches               | 44;  | Conservative 0;  | Mismatches 25; Indels 0; Gaps 0; |
| QY                    | 34   | gaaggtactccacaagcttgaatcccggaagaataagaaacccgagtcctccgacgacag | 93                               |
| Db                    | 245  | GCAGTACTCCACGGGCTTCCCGAGCCCGAACGGGTGACACCGTAGTCCCGCAGCG      | 186                              |
| QY                    | 94   | aacctgcaa 102  |                                  |
| Db                    | 185  | AGCGTCGA 177   |                                  |
| RESULT 11             |  |  |                                  |
| LOCUS                 | BG316445   | 551 bp   | mrna linear EST 26-FEB-2001      |
| DEFINITION            | 947023E01.y1 947 - 2 week shoot from Barkan lab zea mays cDNA, mRNA sequence.  |  |                                  |
| ACCESSION             | BG316445   |  |                                  |
| VERSION               | BG316445.1   | GI:13125875  |                                  |
| KEYWORDS              | EST.   |  |                                  |
| SOURCE                | zea mays.  |  |                                  |
| ORGANISM              | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC   |  |                                  |
| REFERENCE             | 1 (bases 1 to 551)   |  |                                  |
| AUTHORS               | Walbot, V.   |  |                                  |
| TITLE                 | Maize ESTs from various cDNA libraries sequenced at Stanford University  |  |                                  |
| JOURNAL               | Unpublished (1999)   |  |                                  |
| COMMENT               | Contact: Walbot V<br>Department of Biological Sciences<br>Stanford University  |  |                                  |

855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 947023 row: E column: 01

```

FEATURES
source      Location/Qualifiers
           1. .551

```

```

/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="947 - 2 week shoot from Barkan lab"
/tissue_type="leaf and stem, including leaf base"
/dev_stage="2 week old seedling (3 leaves)"
/lab_host="Xil-Blue"
/notes="Organ: shoot; Vector: Lambda ZAP (pluscript SK-);
Site1: EcoRI; Site2: XhoI; Directionally cloned using
Stratagene's Unizap XR cDNA cloning kit with the 5' end
at the EcoRI site. The library represents 8 x 10e5
independent recombinant phage. The plants were greenhouse
grown"

```

| BASE COUNT | 85 a | 197 c | 181 g | 88 t |
|------------|------|-------|-------|------|
| grown.     |      |       |       |      |

|    | Query Match           | 26.13;   | Score 29;       | DB 10;         | Length 551; |
|----|-----------------------|--|-----------------|----------------|-------------|
|    | Best Local Similarity | 63.8%;   | Pred. No. 46;   |                |             |
|    | Matches               | 44;  | Conservative 0; | Mismatches 25; | Indels 0;   |
| Qy | 34                    | gaaggtactccacaagggtttgaatcccggaagaatagaaccgcgagcttcgcgacag     | 93              |                |             |
|    |                       |  |                 |                |             |
| Db | 318                   | GCAGGTACTCCACGGGGTTCCTCCGAGCCGACACGGTTCGAAACCCGTAGTCCGCGACACGC | 259             |                |             |
|    |                       |  |                 |                |             |

Qy 94 aacctgcaa 102  
| | | | |  
Db 258 AGCCGTCGA 250

|  |                                    |
|--|------------------------------------|
| RESULT   | 12                                 |
| BI233858/c   |                                    |
| LOCUS  |                                    |
| DEFINITION   |                                    |
| BI233858   | 566 bp mRNA linear EST 11-JUL-2001 |
| 949037A03.y2 949 - Juvenile leaf and shoot cDNA from Steve Moose |                                    |
| zea mays CDNA, mRNA sequence.                                    |                                    |

zeb mays cdna; mkrna seq  
 BI233858  
 BI233858.1 GI:14701440  
 EST.

SOURCE Zea mays.  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 566)  
 Walbot, V.

| JOURNAL            | TITLE   |
|--------------------|---|
| Unpublished (1999) | Maize ESTs from various cDNA libraries sequenced at Stanford University |

CONFIRMED (1599)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221

FEATURES source  
Location/Qualifiers  
1..566

```

i: 1.300
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moore"
/tissue_type="Immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"

```

```
/lab host="E. coli XL01R"
```

/note="Organ: juvenile vegetative shoots; Vector: PAD-GAU4-2.1; Site\_1: EcoRI; Site\_2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcoRI (5') and XhoI (3') directional cloning into lambda Hybrizap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing - includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developing leaves 4-5 and the vegetative apex, including adult leaf primordia, 9 days after sowing. 4. Partially expanded and greening leaves 4-5 at 13 days after sowing."

|            |                 |       |       |      |
|------------|-----------------|-------|-------|------|
| BASE COUNT | 82 a            | 203 c | 189 g | 92 t |
| ORIGIN     | greening leaves | 4-3   | al    | 13   |

```
Query Match      26.1%; Score 29; DB 10; Length 566;
Best Local Similarity 63.8%; Pred. No. 47;
Matches 44: Conservative 0; Mismatches 25; Indels 0; Gaps 0;
```

| QY | 34  | gaaggtactcacaaagcttgaatcccggaagaatagaaacccgagctctccgacgacag  | 93  |
|----|-----|--|-----|
|    |     |  |     |
| Db | 271 | GCAGGTACTTCCCGAGCGGCTTCCCGAGCGCCGAACCGGTGCAACCGTAGTCGCCGACCA | 212 |
|    |     |  |     |
| QY | 94  | aacctgcaa  | 102 |
|    |     |  |     |
| Db | 211 | AGCCGTCGA  | 203 |

RESULT 13  
TA62H05P

TA62H05P 357 bp DNA linear GSS 16-DEC-2000  
T. brucei sheared genomic DNA clone 62h05, forward sequence,  
genomic survey sequence.

genome survey sequence  
 AL465105  
 AL465105.1 GI:11877561  
 GSS.

| SOURCE              | ORGANISM   |
|---------------------|--|
| Trypanosoma brucei. | Trypanosoma brucei                                       |
| Trypanosoma brucei  | Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; |

Trypanosoma.  
1 (bases 1 to 357)

**AUTHORS**

Hall, N., Bowman, S., Lennard, N. J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S. E., Rajandream, M. A. and Barrell, B. G.

TITLE  
JOURNAL

submitted (10-Dec-2000) trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk

COMMENT

Constructed at the Institute for Genomic Research ("IGR"), Rockville, MD. Genomic DNA isolated from a cloned population of *Trypanosoma brucei* (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 to 4 kb). The vti1 method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford university Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).

|          |        |
|----------|--------|
| FEATURES | SOURCE |
|----------|--------|

```

1.357
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="62h05"
124 a 67 c 76 g 90 t

```

| BASE COUNT | 124 a | 67 c | 76 g | 90 t |
|------------|-------|------|------|------|
| ORIGIN     |       |      |      |      |

```
Query Match      25.9%; Score 28.8; DB 12; Length 357;
Best Local Similarity 54.8%; Pred. No. 50;
Matches 57; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 tctagagcgcaaaaaatcaatcggttactgaggaagtactcccaaggcttgaatcc 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 TATGGGGAAGGACTAATAAGTATAGGACGAAGGGAATATTAGTTGCATTAACTCG 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 cggagaataagaaacgcagctctccgacgacagaacctgcaata 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101 CCACAGAAAAAAAAGTGGCCCGACCAACATCTCTTCACTA 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
LOCUS TA278H10Q 553 bp DNA linear GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 278h10, reverse sequence,
genomic survey sequence.
ACCESSION AL486158
VERSION AL486158.1 GI:11851971
KEYWORDS Trypanosoma brucei.
SOURCE Trypanosoma brucei
ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 553)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
source
1..553
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="278h10"
BASE COUNT 199 a 92 c 99 g 163 t
ORIGIN

Query Match      25.9%; Score 28.8; DB 12; Length 553;
Best Local Similarity 54.8%; Pred. No. 54;
Matches 57; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 tctagagcgcaaaaaatcaatcggttactgaggaagtactcccaaggcttgaatcc 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 TATGGGGAAGGACTAATAAGTATAGGACGAAGGGAATATTAGTTGCATTAACTCG 457
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 cggagaataagaaacgcagctctccgacgacagaacctgcaata 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 CCACAGAAAAAAAAGTGGCCCGACCAACATCTCTTCACTA 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
LOCUS TA174F12Q/c 573 bp DNA linear GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 174f12, reverse sequence,
genomic survey sequence.
ACCESSION AL474540
VERSION AL474540.1 GI:11839754
KEYWORDS Trypanosoma brucei.
SOURCE Trypanosoma brucei
ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 573)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
source
1..573
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="174f12"
BASE COUNT 134 a 123 c 112 g 204 t
ORIGIN

Query Match      25.9%; Score 28.8; DB 12; Length 573;
Best Local Similarity 54.8%; Pred. No. 54;
Matches 57; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 tctagagcgcaaaaaatcaatcggttactgaggaagtactcccaaggcttgaatcc 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 TATGGGGAAGGACTAATAAGTATAGGACGAAGGGAATATTAGTTGCATTAACTCG 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 cggagaataagaaacgcagctctccgacgacagaacctgcaata 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 CCACAGAAAAAAAAGTGGCCCGACCAACATCTCTTCACTA 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: July 15, 2002, 21:48:54
Job time: 18885 sec
```

```
DEFINITION T. brucei sheared genomic DNA clone 174f12, reverse sequence,
genomic survey sequence.
ACCESSION AL474540
VERSION AL474540.1 GI:11839754
KEYWORDS Trypanosoma brucei.
SOURCE Trypanosoma brucei
ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 573)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
source
1..573
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="174f12"
BASE COUNT 134 a 123 c 112 g 204 t
ORIGIN

Query Match      25.9%; Score 28.8; DB 12; Length 573;
Best Local Similarity 54.8%; Pred. No. 54;
Matches 57; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 tctagagcgcaaaaaatcaatcggttactgaggaagtactcccaaggcttgaatcc 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 TATGGGGAAGGACTAATAAGTATAGGACGAAGGGAATATTAGTTGCATTAACTCG 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 cggagaataagaaacgcagctctccgacgacagaacctgcaata 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 CCACAGAAAAAAAAGTGGCCCGACCAACATCTCTTCACTA 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: July 15, 2002, 21:48:54
Job time: 18885 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:28:44 ; Search time 2368.24 Seconds  
(without alignments)  
954.323 Million cell updates/sec

Title: US-10-053-641-4

Perfect score: 108

Sequence: 1 ggcgcctattattgcagg.....aacgcattgatttttttcgc 108

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 1797656 seqs, 10463260293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Query No. | Score | Match | Length | DB | ID | Description |
|------------|-----------|-------|-------|--------|----|----|-------------|
| -----      |           |       |       |        |    |    |             |

|    |      |      |      |    |            |
|----|------|------|------|----|------------|
| 1  | 95.4 | 88.3 | 128  | 6  | I41185     |
| 2  | 95.4 | 88.3 | 235  | 6  | I41188     |
| 3  | 94   | 87.0 | 106  | 6  | I41184     |
| 4  | 94   | 87.0 | 226  | 12 | SYNHIR     |
| 5  | 94   | 87.0 | 235  | 6  | I41187     |
| 6  | 93   | 86.1 | 726  | 6  | I41189     |
| 7  | 93   | 86.1 | 726  | 6  | I41190     |
| 8  | 86.2 | 79.8 | 220  | 12 | SYNHIRMED  |
| 9  | 86.2 | 79.8 | 224  | 6  | A00630     |
| 10 | 86.2 | 79.8 | 224  | 6  | A00631     |
| 11 | 86.2 | 79.8 | 224  | 6  | A04277     |
| 12 | 86.2 | 79.8 | 224  | 6  | A04278     |
| 13 | 86.2 | 79.8 | 224  | 6  | A07345     |
| 14 | 86.2 | 79.8 | 224  | 6  | A07346     |
| 15 | 84.2 | 78.0 | 279  | 6  | A01139     |
| 16 | 84.2 | 78.0 | 279  | 6  | I26634     |
| 17 | 83.2 | 77.0 | 212  | 6  | A13382     |
| 18 | 83.2 | 77.0 | 212  | 6  | A13383     |
| 19 | 83.2 | 77.0 | 212  | 6  | A34619     |
| 20 | 83.2 | 77.0 | 212  | 6  | A34620     |
| 21 | 83.2 | 77.0 | 212  | 6  | E00711     |
| 22 | 83.2 | 77.0 | 8491 | 6  | AR031529   |
| 23 | 82.6 | 76.5 | 122  | 6  | A34236     |
| 24 | 82.6 | 76.5 | 122  | 6  | A34237     |
| 25 | 82.6 | 76.5 | 217  | 6  | A03695     |
| 26 | 82.6 | 76.5 | 217  | 6  | A03696     |
| 27 | 82.6 | 76.5 | 217  | 6  | A34238     |
| 28 | 82.6 | 76.5 | 217  | 6  | A34239     |
| 29 | 82.6 | 76.5 | 217  | 6  | E00637     |
| 30 | 81.4 | 75.4 | 4477 | 12 | SYNEXVECHD |
| 31 | 81.2 | 75.2 | 195  | 6  | A03693     |
| 32 | 81.2 | 75.2 | 195  | 6  | A03694     |
| 33 | 81.2 | 75.2 | 226  | 6  | A04615     |
| 34 | 77.6 | 71.9 | 211  | 6  | E03003     |
| 35 | 76   | 70.4 | 238  | 6  | A04616     |
| 36 | 75.8 | 70.2 | 223  | 6  | A18064     |
| 37 | 75.8 | 70.2 | 223  | 6  | A19994     |
| 38 | 75.8 | 70.2 | 223  | 6  | A19999     |
| 39 | 75.8 | 70.2 | 223  | 6  | A20035     |
| 40 | 75.8 | 70.2 | 223  | 6  | I13188     |
| 41 | 75.8 | 70.2 | 420  | 6  | A20000     |
| 42 | 75.8 | 70.2 | 420  | 6  | I13189     |
| 43 | 75.8 | 70.2 | 1467 | 6  | A20030     |
| 44 | 75.8 | 70.2 | 1467 | 6  | I13218     |
| 45 | 75.4 | 69.8 | 201  | 6  | A34242     |

#### ALIGNMENTS

RESULT 1

| LOCUS      | I41185   | Sequence 5 from patent US 5624822. | 128 bp | DNA | linear | PAT 13-MAY-1997 |
|------------|--|------------------------------------|--------|-----|--------|-----------------|
| DEFINITION | I41185   | Sequence 5 from patent US 5624822. | 128 bp | DNA | linear | PAT 13-MAY-1997 |
| ACCESSION  | I41185   |                                    |        |     |        |                 |
| VERSION    | I41185.1   | GI:2081775                         |        |     |        |                 |
| KEYWORDS   | Unknown.   |                                    |        |     |        |                 |
| SOURCE     | Unknown.   |                                    |        |     |        |                 |
| ORGANISM   | Unknown.   |                                    |        |     |        |                 |
| REFERENCE  | 1 (bases 1 to 128)                                 |                                    |        |     |        |                 |
| AUTHORS    | Koerwer, W.  |                                    |        |     |        |                 |
| TITLE      | Hirudin fusion proteins and preparation of hirudin |                                    |        |     |        |                 |
| JOURNAL    | Patent: US 5624822-A 5 29-APR-1997;                |                                    |        |     |        |                 |
| FEATURES   | Location/Qualifiers                                |                                    |        |     |        |                 |
| source     | 1..128   |                                    |        |     |        |                 |
| BASE COUNT | 20 a 33 c 34 g 41 t                                |                                    |        |     |        |                 |
| ORIGIN     | /organism="unknown"                                |                                    |        |     |        |                 |

Query Match 88.3%; Score 95.4; DB 6; Length 128;  
Best Local Similarity 94.3%; Pred. No. 1.9e-22;





|            |  |
|------------|--|
| RESULT     | 11   |
| A04277/c   |  |
| LOCUS      | A04277   |
| DEFINITION | Artificial sequence for hirudin-like fusion protein. |
| ACCESSION  | A04277   |
| VERSION    | A04277.1   |
| KEYWORDS   | . GI:344867  |
| SOURCE     | synthetic construct.                                 |
| ORGANISM   | synthetic construct.                                 |
|            | linear DNA 224 bp PAT 15-APR-1993                    |

| Query Match   | 79.8%;          | Score 86.2;        | DB 6;     | Length 224; |
|---|-----------------|--------------------|-----------|-------------|
| Best Local Similarity   | 87.9%;          | Pred. No. 3.3e-19; |           |             |
| Matches 94;   | Conservative 0; | Mismatches 13;     | Indels 0; | Gaps 0;     |
| 2 gccgcgcctattatgcaggtattcttcocgggattcttcaagtcgcgcgtgttgaga 61    |                 |                    |           |             |
|   |                 |                    |           |             |
| 9 GCTCACTTACTATTGCAGGTATTTCTTCGSGGATTCTTCGAAGTCACCGTCGTTGTGAGA 68 |                 |                    |           |             |
| 62 ctgcggtttcggagctaccttcgcagtaacgcgaattgatttttttcgc 108          |                 |                    |           |             |
| 69 CTGCGGTTTCGGAGTACTCTCACGGGTACGCACTGGTCTCTTTCTTAC 115           |                 |                    |           |             |

|            |  |
|------------|--|
| RESULT     | 13   |
| A07345/c   |  |
| LOCUS      | A07345                                     |
| DEFINITION | Synthetic DNA for hirudin-similar protein. |
| ACCESSION  | A07345                                     |
| VERSION    | A07345.1                                   |
| KEYWORDS   | . GI:413049                                |
| SOURCE     | synthetic construct.                       |
| ORGANISM   | synthetic construct                        |
|            | linear                                     |
|            | 224 bp                                     |
|            | DNA  |
|            | PAT 03-AUG-1993                            |







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:48 ; Search time 758.37 seconds  
(without alignments)  
244.507 Million cell updates/sec

Title: US-10-053-641-4

Perfect score: 108

Sequence: 1 ggcgcctattattgcagg.....aacgcattgatttttttcgc 108

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

- 1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | ID     | Description |
|------------|-------|-------------|--------|-------------|
| C 1        | 93    | 86.1        | 726 12 | AAQ12379    |
| C 2        | 85.4  | 79.1        | 224 7  | AAQ12379    |
| C 3        | 84.2  | 78.0        | 279 12 | AAQ13831    |
| C 4        | 83.8  | 77.6        | 468 16 | AAQ93244    |
| C 5        | 83.2  | 77.0        | 227 22 | AAQ61507    |
| C 6        | 82.6  | 76.5        | 217 7  | AAQ60355    |
| C 7        | 82.6  | 76.5        | 217 8  | AAQ70323    |
| C 8        | 82.4  | 76.3        | 185 12 | AAQ14926    |
| C 9        | 77.6  | 71.9        | 210 12 | AAQ10452    |

Hirudin peptide/Pr  
DNA encoding hirud  
MSP signal peptide  
Fusion Construct o  
S. marcescens hiru  
Desulphatohirudin  
Sequence of the de  
Synthetic hirudin  
HV-1 gene. Synthe

Sequence encoding  
Factor Xa-cleavabl  
Factor Xa-cleavabl  
Desulphatohirudin  
Desulphatohirudin  
Synthetic hirudin  
Hirudin HV-1. Syn  
Desulphatohirudin  
Synthetic HVI gene  
Synthetic HVI gene  
Synthetic hirudin  
CUPI promoter, PHO  
Yeast CUP1 promote  
pJDB207/GAPFL-YHIR  
Leech hirudin mute  
Sequence encoding  
Sequence encoding  
Factor Xa-cleavabl  
Encodes hirudin de  
Sequence of DNA in  
Encodes hirudin de  
Secretion plasmid  
Olesein-hirudin fu  
Desulphatohirudin  
Sequence encoding  
Hirudin variant.  
Partial Hirudin HV  
DNA encoding the f  
rhv2-Phe3.Gln33,Ty  
DNA encoding hirud  
Sequence encoding  
Hirudin gene HVI.  
Sequence encoding  
Recombinant hirudi  
DNA encoding leech  
DNA encoding leech

#### ALIGNMENTS

RESULT 1  
AAQ12379/c  
ID AAQ12379 standard; DNA; 726 BP.  
AC AAQ12379;  
XX  
XX  
DT 17-SEP-1991 (first entry)  
XX  
DE Hirudin peptide/Protein A fusion gene.  
XX  
KW anticoagulant; fusion protein; ds.  
FH Key Location/Qualifiers  
FT mat\_peptide 1..489  
FT FT /\*tag= a  
FT FT /product= Protein A  
FT mat\_peptide 529..720  
FT FT /\*tag= b  
FT FT /product= Hirudin  
FT FT 490..528  
FT FT /\*tag= c  
FT FT /product= linker oligopeptide Y  
XX  
PN DE3942580-A.  
XX  
PD 27-JUN-1991.  
XX  
PF 22-DEC-1989; 89DE-3942580.  
XX  
PR 22-DEC-1989; 89DE-3942580.  
XX  
PA (BADI ) BASF AG.

XX PI Korwer W;  
 XX DR WPI; 1991-194236/27.  
 XX DR P-PSDB; AAR12751.  
 XX PT Hirudin peptide prodn. by cleaving new fusion peptide - of  
 PT hirudin and protein A, expressed in high yield and stable,  
 PT soluble form by transformed E.coli  
 XX PS Example 1; Page 6-7; 9pp; German.  
 XX CC This sequence is an example of a fusion construct for expression of  
 CC the fusion peptide of the invention. The Met residue in the linker  
 CC oligopeptide allows cleavage by CNBr to release two fragments which  
 CC can be easily separated by Igg affinity chromatography. (The protein  
 CC A component binds to Igg sepharose). Increased yields of hirudin are  
 CC obtained using this fusion construct.  
 XX CC  
 XX SQ Sequence 726 BP; 280 A; 171 C; 126 G; 149 T; 0 other;

Query Match 86.1%; Score 93; DB 12; Length 726;  
 Best Local Similarity 95.0%; Pred. No. 2.3e-22;  
 Matches 96; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 ctattattgcaggtattcttcgggatttctcaaaagtcgcgtgtgtgagactgcgg 67  
 Db TTTTACTGTCAGGTATTCTTCGGGATTTCTTCGAGTCGCGTGTGTGAGACTGCGG 667

QY 68 ttccggagtagcttcgcagtaacgcattgattttttcgc 108  
 Db TTTTCGGGTACTTCGCCAGTAACGCACACTGGTTTTTTTCGC 526

RESULT 2  
 AAN60746/C  
 ID AAN60746 standard; DNA; 224 BP.  
 XX AC AAN60746;  
 XX DT 01-JAN-1980 (first entry)  
 XX DE DNA encoding hirudin-like polypeptide.  
 XX KW Hirudin; thrombin-antagonist; anticoagulant; ds.  
 XX OS Synthetic.  
 XX PN DE3445517-A.  
 XX PD 19-JUN-1986.  
 XX PF 13-DEC-1984; 84DE-3445517.  
 XX PR 13-DEC-1984; 84DE-3445517.  
 XX PA (GENB-) GEN-BIO-TEC GES GEN.  
 XX PI Fortkamp E, Rieger M, Sommer R;  
 XX DR WPI; 1986-162802/26.  
 XX DR P-PSDB; AAP60827.  
 XX CC New DNA sequence coding for new hirudin like polypeptide - useful  
 PT as thrombin antagonist, e.g. for inhibiting blood coagulation  
 XX PS Disclosure; Fig. 1; 25pp; German.  
 XX CC The sequence encodes a protein with hirudin-like activity. It is a  
 CC thrombin-antagonist useful for inhibiting blood coagulation and for  
 CC treating inflammation and/or oedema. It can be prepared in large  
 CC amounts, unlike natural hirudin which is available only in tiny

CC amounts from leeches. This polypeptide is produced using  
 CC recombinant DNA techniques and is expressed in a bacterial  
 XX transformant e.g. Escherichia coli.  
 XX SQ Sequence 224 BP; 61 A; 51 C; 60 G; 51 T; 1 other;

Query Match 79.1%; Score 85.4; DB 7; Length 224;  
 Best Local Similarity 86.9%; Pred. No. 6.9e-20;  
 Matches 93; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

QY 2 gccgcctattattgcaggtattcttcgggatttctcaaaagtcgcgtgtgtgaga 61  
 Db TTTTACTGTCAGGTATTCTTCGGGATTTCTTCGAGTCGCGTGTGTGAGACTGCGG 157

QY 62 ctgcgggttcgcagtagcttcgcagtaacgcattgattttttcgc 108  
 Db TTTTCGGGTACTTCGCCAGTAACGCACACTGGTTTTTTTCAC 110

RESULT 3  
 AAQ13831/C  
 ID AAQ13831 standard; DNA; 279 BP.  
 XX AC AAQ13831;  
 XX DT 09-DEC-1991 (first entry)  
 XX DE MSP signal peptide-hirudin gene fusion.  
 XX KW Major secretion product; expression cassette; desulfatohirudin; ss.  
 XX OS Lactococcus lactis LM0230 (DSM 5805), Hirudo medicinalis.  
 XX FH Key Location/Qualifiers  
 FT sig\_peptide 1..81  
 FT /tag= a  
 FT /note= "MSP signal peptide"  
 FT mat\_peptide 82..279  
 FT /tag= b  
 FT /note= "hirudin structural gene."  
 XX PN EP449770-A.  
 XX PD 02-OCT-1991.  
 XX PF 13-MAR-1991; 91EP-0810167.  
 XX PR 22-MAR-1990; 90GB-0006400.  
 XX PA (CIBA ) CIBA GEIGY AG.  
 XX PI Suri B, Schmitz A;  
 XX DR WPI; 1991-290162/40.  
 XX DR P-PSDB; AAR14151.  
 XX PT Hybrid vectors for expression of polypeptide(s) - comprise DNA  
 PT from plasmids obt'd. from Lactococcus lactis, esp. major secretion  
 PT prod. gene fragments.  
 XX PS Disclosure; Page 26; 36pp; English.  
 XX CC The sequence (SEQ ID NO:2) was obt'd. by ligation of the signal  
 CC peptide encoding DNA from L. lactis (from pUCRS, DSM 5803) and the  
 CC coding region for desulfatohirudin (from pML310, EP-168342).  
 CC The presence of the signal peptide results in the secretion of  
 CC expressed hirudin into the supernatant.  
 CC See also AAQ13830.  
 XX SQ Sequence 279 BP; 76 A; 66 C; 68 G; 69 T; 0 other;



CC with the Leu-hirudin (LH) ((Leu1-Thr2)-63-desulfato-hirudin) sequence  
 CC linked to the C-terminus of the signal sequence. (I) is an intermediate  
 CC in recombinant production of LH, a known antithrombotic. The specified  
 CC signal sequence may also be used for secretory expression of other  
 CC proteins. (II) is processed directly to LH and this, in native form,  
 CC secreted from E. coli in high yield. This results, both during  
 CC fermentation and subsequent purification, in a higher concentration of  
 CC hirudin, reducing costs of production. The specified signal sequences  
 CC provide more efficient secretion than known sequences. This sequence  
 CC encodes a fragment of the S. marcescens hirudin protein.  
 XX  
 SQ Sequence 227 BP; 61 A; 60 C; 58 G; 48 T; 0 other;

Query Match 77.0%; Score 83.2; DB 22; Length 227;  
 Best Local Similarity 87.5%; Pred. No. 4e-19;  
 Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 5 gccctattatgcaggtattcttcggtattcttcaagtcgcgcgtgtgtgagactg 64  
 DB 204 GCTCTATTACTGAAGGTATTCTCTAGGATCTCTTCGAAGTCGCGGTATGAGACTG 145  
 QY 65 cggttcggagtaccttcgcagtaacgcattgatttttttcgc 108  
 DB 144 CGGTTTCGGGTACCTTCGCAGTAACGCACCTGCTCTTTTCAC 101

RESULT 6  
 AAN60355/c  
 ID AAN60355 standard; DNA; 217 BP.  
 XX  
 AC AAN60355;  
 XX  
 DT 20-JUN-1991 (first entry)  
 XX  
 DE Desulphatohirudin (II).  
 XX  
 KW Desulphatohirudin; antibodies; thrombin; ds.  
 XX  
 OS Synthetic.

Key Location/Qualifiers  
 CDS 10..207  
 FT /\*tag= a  
 FT /product= hirudin  
 XX  
 PN EP168342-A.

XX  
 PD 15-JAN-1986.  
 XX  
 PF 10-JUN-1985; 85EP-0810268.  
 XX  
 PR 14-JUN-1984; 84CH-0288284.  
 XX  
 PA (CIBA ) CIBA GEIGY AG.  
 XX  
 PL Liersch M, Rink H, Marki W, Grutter MG, Meyhack B;  
 XX  
 WPI; 1986-015589/03.  
 DR P-PSDB; AAP60395.  
 XX

DNA sequences coding for hirudin and derivs. - and expression  
 PT vectors, transformed cells, monoclonal antibodies and hybridomas,  
 PT useful as thrombin inhibitor.

PS Disclosure; Page 10; 123pp; German.

XX The sequence comprises an EcoRI restriction enzyme site at the  
 CC 5'-end and a BamHI restriction enzyme site at the 3'-end.  
 CC The sequence may be introduced into a vector for the transformation  
 CC of hosts, e.g. E. coli. The hirudin can thus be prepared on a  
 CC large scale. The product and its derivs. are thrombin inhibitors,  
 CC useful in anticoagulant therapy, esp. when injected at doses

CC of 0.01-0.05 mg/kg.

XX  
 SQ Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;

Query Match 76.5%; Score 82.6; DB 7; Length 217;  
 Best Local Similarity 90.7%; Pred. No. 6.3e-19;  
 Matches 88; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 12 tattgcaggtattcttcggtattcttcaagtcgcgcgtgtgtgagactgcggttc 71  
 DB 209 TACTGCAGGTATTCTTCGGGATTTCTTCGAAGTCACCGTCGTTGTGAGACTCGGTTTC 150  
 QY 72 ggagtaccttcgcagtaacgcattgatttttttcgc 108  
 DB 149 GGGTACCTTCACCGGTAAAGCACTGGTTTTTTTCAC 113

RESULT 7  
 AAN70323/c  
 ID AAN70323 standard; DNA; 217 BP.  
 XX  
 AC AAN70323;  
 XX  
 DT 02-APR-1991 (first entry)  
 XX  
 DE Sequence of the desulphatohirudin gene.  
 XX  
 KW Anticoagulant; thrombin inhibitor; ds.  
 XX  
 PN EP225633-A.  
 XX  
 PD 16-JUN-1987.  
 XX  
 PF 09-DEC-1986; 86EP-0117098.  
 XX  
 PR 29-MAY-1986; 86GB-0013088.  
 PR 12-DEC-1985; 85GB-0030631.  
 XX  
 PA (CIBA ) CIBA GEIGY AG.  
 PA (PLAN-) PLANTORGAN WERK HEINRICH.  
 PA (CHRI-) PLANTORGANW CHRISTENSEN.  
 XX  
 PI Meyhack B, Marki W, Heim J;  
 XX  
 WPI; 1987-164868/24.

XX New DNA constructs and hybrid vectors for transformation of yeast  
 PT etc. - useful for prodn. and secretion of protein with hirudin  
 PT activity for use as thrombin inhibitors.  
 XX  
 PS Example; p44; 146pp; English.  
 XX  
 CC The preferred DNA construct of the invention contains the PHO5  
 CC promoter and a DNA segment consisting of the PHO5 signal sequence  
 CC upstream of and in reading frame with a DNA sequence coding for  
 CC mature desulphatohirudin. The segment is under the transcriptional  
 CC control of the PHO5 promoter and the 3' flanking sequence of the  
 CC PHO5 gene.  
 XX  
 SQ Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;

Query Match 76.5%; Score 82.6; DB 8; Length 217;  
 Best Local Similarity 90.7%; Pred. No. 6.3e-19;  
 Matches 88; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 12 tattgcaggtattcttcggtattcttcaagtcgcgcgtgtgtgagactgcggttc 71  
 DB 209 TACTGCAGGTATTCTTCGGGATTTCTTCGAAGTCACCGTCGTTGTGAGACTCGGTTTC 150  
 QY 72 ggagtaccttcgcagtaacgcattgatttttttcgc 108  
 DB 149 GGGTACCTTCACCGGTAAAGCACTGGTTTTTTTCAC 113



```

FH Key          Location/Qualifiers
FT CDS          1..237
FT             /*tag= a
XX
PN EP252854-A.
XX
XX 13-JAN-1988.
XX
XX 10-JUL-1987; 87EP-0401649.
XX
XX 01-DEC-1986; 86FR-0016722.
XX
XX 01-JUL-1986; 86FR-0010090.
XX
XX (TRAN-) TRANSCENE SA.
XX
XX Labat N, Loison G, Bolland A, Lemoine Y;
XX
XX WPI; 1988-008792/02.
XX
XX P-PSDB; AAP80995.
XX
XX New DNA block providing expression of hirudin in yeast -
XX comprises hirudin gene, signal and leader sequences and sequence
XX for unique proteolytic cleavage site
XX
XX Disclosure; ; pp; French.
XX
XX The patent is for a new DNA block providing expression of hirudin (H) in
XX yeast. Also new are plasmids contg. the block plus at least one yeast
XX origin of replication and yeasts, esp. Saccharomyces cerevisiae,
XX transformed with such plasmids.
XX
XX Sequence 238 BP; 79 A; 47 C; 55 G; 57 T; 0 other;

Query Match          70.4%; Score 76; DB 9; Length 238;
Best Local Similarity 85.0%; Pred. No. 1.2e-16;
Matches 85; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 9 tattattgcaggtattcttcgggatttcttcaaaagtcgcgcgtgtgtgagactcggt 68
Db 203 TTTTCATGTAAATATCTTCTGGAAATTTCTCGAAGTCACCGTCGTTATGAGACTGCGGT 144

Qy 69 ttcggagtagcttcgcgaagtaacgcattgattttttcgc 108
Db 143 TTCGGAGTACTTTCACCTGTAAACGCACTGTTTTCAC 104

RESULT 11
AAQ12155/c
XX AAQ12155 standard; DNA; 420 BP.
XX
XX AAQ12155;
XX
XX 17-SEP-1991 (first entry)
XX
XX Factor Xa-cleavable hirudin-IEGR-hirudin gene.
XX
XX Fusion protein; blood clotting; coagulation; fibrinolysis;
XX antithrombotic; thrombolysis; streptokinase; ss.
XX
XX Synthetic.
XX
XX Key          Location/Qualifiers
XX CDS          1..195
XX             /*tag= a
XX             /label= hirudin gene
XX             196...207
XX             /*tag= b
XX             /label= linker
XX             /note= "encodes cleavage site IEGR"
XX             208...408
XX             /*tag= c
XX             /label= hirudin gene

```

```

XX WO9109125-A.
XX
XX 27-JUN-1991.
XX
XX 07-DEC-1990; 90WO-GH01911.
XX
XX 07-DEC-1990; 90WO-GH01911.
XX
XX 07-DEC-1989; 89GB-0027722.
XX
XX (BRBI-) BRIT BIO-TECHN LTD.
XX
XX Dawson KM, Hunter MG, Czaplewski LG;
XX
XX WPI; 1991-208151/28.
XX
XX P-PSDB; AAR12888.
XX
XX Fusion protein cleavage by blood clotting enzyme - for prodn. of
XX fractions having greater antithrombotic activity for therapy and
XX prophylaxis.
XX
XX Disclosure; Page 78; 115pp; English.
XX
XX The sequence of the synthetic hirudin HV-1 genes was designed
XX based on the published amino acid sequence (Dodd J., et al FEBS
XX Letters 165 180 (1984)). Unique restriction sites were incorpor-
XX ated to facilitate subsequent genetic manipulation. The codons
XX selected were those favourable for S. cerevisiae or E. coli. The
XX sequence was divided into 12 oligomers which were synthesised and
XX then annealed. The ligation prod. was ligated to HindIII and EcoRI
XX treated pUC19 plasmid DNA and the resulting vector used to trans-
XX form E. coli K12 HW87. Plasmid pUC19 HV-1 was isolated from trans-
XX formants and inserted into plasmid pSW6, a shuttle vector, for ex-
XX pression. The gene was then used to construct an expression vector
XX in which two hirudin genes are linked together via a linker encod-
XX ing a cleavage site for factor Xa. The factor Xa is present at the
XX site of the target thrombus so the active agents are released
XX specifically at the place where clot formation is occurring.
XX See also AAQ12153-Q12156, AAQ12158-Q12162 and AAQ12490.
XX
XX Sequence 420 BP; 124 A; 90 C; 110 G; 96 T; 0 other;

Query Match          70.2%; Score 75.8; DB 12; Length 420;
Best Local Similarity 83.5%; Pred. No. 1.7e-16;
Matches 86; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 6 cccattattgcaggtattcttcgggatttcttcaaaagtcgcgcgtgtgtgagactgc 65
Db 410 CCTATTACTGCAGATATTTCTTGGGATTTCTTGGAAATCTCCATCTGTGGACTGC 351

Qy 66 gggttcggagtagcttcgcgaagcagcattgattttttcgc 108
Db 350 GGCCTTTGGGTACTTTCACCAGTGACACATTTGGTCTTTTCAC 308

RESULT 12
AAQ12490/c
XX AAQ12490 standard; DNA; 1467 BP.
XX
XX AAQ12490;
XX
XX 17-SEP-1991 (first entry)
XX
XX Factor Xa-cleavable streptokinase-IEGR-hirudin gene.
XX
XX Fusion protein; blood clotting; coagulation; fibrinolysis;
XX antithrombotic; thrombolysis; ss.
XX
XX Synthetic.
XX
XX Key          Location/Qualifiers
XX CDS          1..1455

```



```

FT mat_peptide /*tag= a
FT 1..1242
FT /*tag= b
FT /label= streptokinase
FT 1243..1254
FT /*tag= c
FT /label= linker
FT /note= "encodes factor Xa cleavage site"
FT mat_peptide 1255..1453
FT /*tag= d
FT /label= hirudin HV-1
XX
XX WO9109125-A.
XX
XX 27-JUN-1991.
XX
XX 07-DEC-1990; 90WO-GB01911.
XX
XX 07-DEC-1990; 90WO-GB01911.
XX 07-DEC-1989; 89GB-0027722.
XX
XX (BRBI-) BRIT BIO-TECHN LTD.
XX
XX Dawson KM, Hunter MG, Czapleswski LG;
XX
XX WPI: 1991-208151/28.
XX P-PSDB; AARI2522.
XX
XX Fusion protein cleavage by blood clotting enzyme - for prodn. of
XX fractions having greater antithrombotic activity for therapy and
XX prophylaxis.
XX
XX Disclosure; Page 98; 115pp; English.
XX
XX The sequence of the synthetic hirudin HV-1 gene was designed
XX based on the published amino acid sequence (Dodd J., et al FEBS
XX Letters 165 180 (1984)). The sequence of streptokinase was obtd.
XX from PCR amplified chromosomal DNA from S. equisimilis ATCC 10009
XX or ATCC 9642. The primers used for the PCR were based on the pub-
XX lished DNA sequence of S. equisimilis strain H46A (Malke, H., Roe,
XX B., and Ferretti, J.J., Gene 34 357-362 (1985)). The two
XX sequences were used to construct an expression vector in which the
XX streptokinase gene is linked to the hirudin gene via a linking
XX sequence encoding a cleavage site for factor Xa. The factor Xa is
XX present at the site of the target thrombus so the active agents are
XX released specifically at the place where clot formation is occurring.
XX See also AAQ12153-Q12156 and AAQ12158-Q12162.
XX
XX Sequence 1467 BP; 494 A; 317 C; 292 G; 364 T; 0 other;

Query Match 70.2%; Score 75.8; DB 12; Length 1467;
Best Local Similarity 83.5%; Pred. No. 2.3e-16;
Matches 86; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 6 cccattatgcaggattcttcgggattcttcgaagtcgcgcgtgttgtagactgc 65
DB 1457 CCCTATTACGTCAGATATTCTTCGGGATTTCTCGAAATCTCCATCGTTGTGGACTGC 1398

QY 66 ggttcggagtagcttcgccagtaaacgcatgttttttcgc 108
DB 1397 GGCTTTGGGTACCTTCACCAAGTCACACATGGTCTTTTCAC 1355

RESULT 13
AAQ54997/c
ID AAQ54997 standard; DNA; 567 BP.
XX
XX AAQ54997;
XX
XX 11-JUL-1994 (first entry)
XX
XX Desulphatohirudin gene expression/excretion cassette.

```

```

XX
XX Hirudin; HV-1; E. coli; codon usage; desulphatohirudin; HV01; 33Aasp;
XX 33 Asn; biological activity; thrombosis; thromboembolism; ss.
XX
XX Hirudo medicinalis.
XX
XX Key Location/Qualifiers
XX -35_signal 43..48
XX /*tag= a
XX -10_signal 68..73
XX /*tag= b
XX misc_signal 83
XX /*tag= c
XX /note= "Transcription Initiation site"
XX RBS 260..266
XX /*tag= d
XX /note= "Shine-Delgarno sequence"
XX sig_peptide 269..352
XX /*tag= e
XX mat_peptide 353..550
XX /*tag= f
XX /product= Hirudin
XX
XX EP576792-A.
XX
XX 05-JAN-1994.
XX
XX 13-APR-1993; 93EP-0105848.
XX
XX 09-APR-1992; 92HU-0001200.
XX
XX (BIOG ) BIOGAL GYOGYSZERGYAR.
XX
XX Albrecht K, Ambrus G, Barabas E, Barta I, Berk T;
XX Botond K, Egyued C, Erdei J, Kiss P, Klupp T, Koenczoel K;
XX Mate G, Molnari, Moravcsik I, Ott I, Patthy A;
XX Poelya K, Salat J, Tegdes A, Vincze A, Zilahi E;
XX
XX WPI: 1994-009153/02.
XX P-PSDB; AAR47490.
XX
XX Prodn. of recombinant desulphatohirudin HV-1 peptide(s) - using
XX E. coli, Saccharomyces and Streptomyces hosts, for increased
XX yields
XX
XX Disclosure; Page 48; 79pp; English.
XX
XX This sequence represents the an expression/secretion cassette for
XX the expression of hirudin HV-1. This sequence was used in the
XX production of desulphatohirudin HV01 33Aasp and desulphatohirudin HV-1
XX 33 Asn. The expressed peptides produced using this sequence have the
XX same biological activity as natural hirudin and can be used in the
XX treatment of thrombosis, thromboembolism, etc. Using naturalised
XX sequences such as this, large amounts of hirudin may be produced
XX stably, with the highest production level achieved being 140-180
XX mg/litre of culture.
XX
XX Sequence 567 BP; 162 A; 123 C; 134 G; 148 T; 0 other;

Query Match 69.4%; Score 75; DB 15; Length 567;
Best Local Similarity 81.3%; Pred. No. 3.4e-16;
Matches 87; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 2 gccgcctattattgcaggattcttcgggattcttcgaagtcgcgcgtgttgtaga 61
DB 559 GCATGCTTACTATTGTAGGTATTTCTTCAGGAATTTCTCGAAGTCACCATCTGTTGGGA 500

QY 62 ctgcgggttcggagtagcttcgcagtaacgcattgatttttttcgc 108
DB 499 TTGTGGCTTTGGAGTACCTTCGCCAGTGACACATTGATTTTTTTCAC 453

```

RESULT 14  
AAQ54995/c  
ID AAQ54995 standard; DNA; 219 BP.  
XX AC AAQ54995;  
XX AC AAQ54995;  
XX 11-JUL-1994 (first entry)  
XX Desulphatohirudin gene with E. coli-Saccharomyces codon usage.  
DE Hirudin; HV-1; E. coli; codon usage; desulphatohirudin; HV01; 33Asp;  
KW 33 Asn; biological activity; thrombosis; thromboembolism; ds.  
XX Hirudo medicinalis.  
OS  
XX Key Location/Qualifiers  
FH 14..208  
FT CDS /\*tag= a  
FT /product= Hirudin\_HV-1  
XX EP576792-A.  
XX 05-JAN-1994.  
XX 13-APR-1993; 93EP-0105848.  
XX 09-APR-1992; 92HU-0001200.  
XX (BIOG ) BIOGAL GYOGYSZERGYAR.  
XX Albrecht K, Ambrus G, Barabas E, Barta I, Berk T;  
PI Bocond K, Egyued C, Erdei J, Kiss P, Klupp T, Koenczoel K;  
PI Mace G, Molnari, Moravcsik I, Ott I, Patthy A;  
PI Poelya K, Salat J, Tegdes A, Vincze A, Zillahi E;  
XX WPI: 1994-009153/02.  
DR P-PSDB; AAR47488.  
XX  
XX Prodn. of recombinant desulphatohirudin HV-1 peptide(s) - using  
PT E. coli, Saccharomyces and Streptomyces hosts, for increased  
PT yields  
XX Disclosure; Page 44; 79pp; English.  
XX This sequence represents the hirudin HV-1 gene designed on the  
CC basis on E. coli codon usage. This sequence was used in the  
CC production of desulphatohirudin HV01 33Asp and desulphatohirudin HV-1  
CC 33 Asn. The expressed peptides produced using this sequence have the  
CC same biological activity as natural hirudin and can be used in the  
CC treatment of thrombosis, thromboembolism, etc. Using naturalised  
CC sequences such as this, large amounts of hirudin may be produced  
CC stably, with the highest production level achieved being 140-180  
CC mg/litre of culture.  
XX  
SQ Sequence 219 BP; 63 A; 46 C; 51 G; 58 T; 1 other;

Query Match 69.3%; Score 74.8; DB 15; Length 219;  
Best Local Similarity 83.3%; Pred. No. 3.le-16;  
Matches 85; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 7 cctattattgcaggattcttcgggatttcttcaaaagtcgcgcgtgtgtgagactgog 66  
DB 215 CTATTATTGTAGTATTCTTCAGGAATTCCTCGAAGTCACCATCTGTGGGATTGTG 156  
QY 67 gtttcgagtagcttcctccagtaacgcattgatttttttcgc 108  
DB 155 GCTTTGGAGTACCTTCGCCAGTGACACATTGATTTTTCAC 114

RESULT 15  
AAQ12153/c  
ID AAQ12153 standard; DNA; 201 BP.

XX AAQ12153;  
XX 17-SEP-1991 (first entry)  
XX Synthetic hirudin type HV-1 gene.  
XX Fusion protein; blood clotting; coagulation; fibrinolysis;  
KW antithrombotic; thrombolysis; streptokinase; ds.  
XX Synthetic.  
XX Key Location/Qualifiers  
FH 1..201  
FT CDS /\*tag= a  
FT  
XX WO9109125-A.  
XX 27-JUN-1991.  
XX 07-DEC-1990; 90WO-GB01911.  
XX 07-DEC-1990; 90WO-GB01911.  
XX 07-DEC-1989; 89GB-0027722.  
XX (BRBI-) BRIT BIO-TECHN LTD.  
XX Dawson KM, Hunter MG, Czaplowski LG;  
XX WPI: 1991-208151/28.  
DR P-PSDB; AAR12887.  
XX Fusion protein cleavage by blood clotting enzyme - for prodn. of  
PT fractions having greater antithrombotic activity for therapy and  
PT prophylaxis  
XX Disclosure; Page 68; 115pp; English.  
XX The sequence was designed based on the published amino acid se-  
CC quence (Dodt J., et al FEBS Letters 165 180 (1984)). Unique re-  
CC striction sites were incorporated to facilitate subsequent genetic  
CC manipulation. The codons selected were those favourable for S.  
CC cerevisiae or E. coli. The sequence was divided into 12 oligomers  
CC which were synthesised and then annealed. The ligation prod. was  
CC ligated to HindIII and EcoRI treated pUC19 plasmid DNA and the  
CC resulting vector used to transform E. coli K12 HB87. Plasmid pUC19  
CC HV-1 was isolated from transformants and inserted into plasmid  
CC pSW6, a shuttle vector, for expression. The gene can be used to  
CC construct expression vectors in which the hirudin gene is linked to  
CC a second gene encoding e.g. another hirudin protein, streptokinase  
CC or a streptokinase-like protein, via a linking peptide. This pep-  
CC tide link contains a cleavage site for e.g. factor X or thrombin  
CC which can be cleaved, releasing the individual proteins which have  
CC antithrombotic activity. The enzymes which cleave the fusion pro-  
CC tein are present at the site of the target thrombus so the active  
CC agents are released specifically at the place where clot formation  
CC is occurring.  
CC See also AAQ12154-Q12156, AAQ12158-Q12162 and AAQ12490.

Sequence 201 BP; 59 A; 43 C; 52 G; 47 T; 0 other;

Query Match 68.3%; Score 73.8; DB 12; Length 201;  
Best Local Similarity 83.2%; Pred. No. 6.7e-16;  
Matches 84; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 8 ctattattgcaggattcttcgggatttcttcaaaagtcgcgcgtgtgtgagactgog 67  
DB 201 CTATTACTGCAGATATCTCTGGGATTTCTCGAATCTCCATCTCGTGTGGAGTCGGG 142  
QY 68 ttctcgagtagcttcctccagtaacgcattgatttttttcgc 108  
DB 141 CTTTGGGGTACCTTCACAGTGCACACATTGGTCTTTTCAC 101

Tue Jul 16 08:24:08 2002

Search completed: July 15, 2002, 22:44:49  
Job time: 10062 sec

us-10-053-641-4.rng

---

Page 9







APPLICATION NUMBER: US 07861820  
FILING DATE: 18-JUN-1992  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-262-384A-7

Query Match 87.0%; Score 94; DB 1; Length 235;  
Best Local Similarity 95.1%; Pred. No. 1.1e-23;  
Matches 97; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 cctattatgcaggtattcttcggtattcttcaagtcgcccgtgtgtgagactgag 66  
DB 235 CCTATTACTGCAGGTATTCTTCGGGATTCTTCGAAGTCGCGCTGTGTGAGACTGG 176  
QY 67 gtttcgaggtaccttcgagtaacgcatgtatttttcgc 108  
DB 175 GTTTCGGGTACCTTCGCCAGTAACGCACTGGTTTTTTTCGC 134

## RESULT 5

US-08-262-384A-9/c  
Sequence 9, Application US/08262384A  
Patent No. 5624822

GENERAL INFORMATION:  
APPLICANT: Koerwer, Wolfgang  
TITLE OF INVENTION: The Preparation of Hirudin  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kell & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM AT-compatible, 80486 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/262,384A  
FILING DATE: 20-JUN-1994

CLASSIFICATION: 530  
CLASSIFICATION: C 12 N 15/62  
CLASSIFICATION: C 12 N 15/31  
CLASSIFICATION: C 07 K 7/10  
CLASSIFICATION: C 12 P 21/02

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP90/02084  
FILING DATE: 04-DEC-1990  
APPLICATION NUMBER: US 07861820

FILING DATE: 18-JUN-1992  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

US-08-262-384A-9

Query Match 86.1%; Score 93; DB 1; Length 726;  
Best Local Similarity 95.0%; Pred. No. 3.5e-23;  
Matches 96; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 ctattattgcaggtattcttcggtattcttcaagtcgcccgtgtgtgagactgag 67  
DB 726 CTATTACTGCAGGTATTCTTCGGGATTCTTCGAAGTCGCGCTGTGTGAGACTGG 667

QY 68 ttctcgaggtaccttcgagtaacgcatgtatttttcgc 108  
DB 666 TTTTCGGGTACCTTCGCCAGTAACGCACTGGTTTTTTTCGC 626

## RESULT 6

US-08-262-384A-10  
Sequence 10, Application US/08262384A  
Patent No. 5624822

GENERAL INFORMATION:  
APPLICANT: Koerwer, Wolfgang  
TITLE OF INVENTION: The Preparation of Hirudin  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kell & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM AT-compatible, 80486 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/262,384A  
FILING DATE: 20-JUN-1994

CLASSIFICATION: 530  
CLASSIFICATION: C 12 N 15/62  
CLASSIFICATION: C 12 N 15/31  
CLASSIFICATION: C 07 K 7/10  
CLASSIFICATION: C 12 P 21/02

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP90/02084  
FILING DATE: 04-DEC-1990  
APPLICATION NUMBER: US 07861820

FILING DATE: 18-JUN-1992  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

US-08-262-384A-10

Query Match 86.1%; Score 93; DB 1; Length 726;  
Best Local Similarity 95.0%; Pred. No. 3.5e-23;  
Matches 96; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 ctattattgcaggtattcttcggtattcttcaagtcgcccgtgtgtgagactgag 67  
DB 1 CTATTACTGCAGGTATTCTTCGGGATTCTTCGAAGTCGCGCTGTGTGAGACTGG 60

QY 68 ttctcgaggtaccttcgagtaacgcatgtatttttcgc 108  
DB 61 TTTTCGGGTACCTTCGCCAGTAACGCACTGGTTTTTTTCGC 101

## RESULT 7

US-08-186-222-3/c  
Sequence 3, Application US/08186222  
Patent No. 5559007

GENERAL INFORMATION:  
APPLICANT: Suri, Bruno  
APPLICANT: Schmitz, Albert  
TITLE OF INVENTION: Bacterial Vectors  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive

CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/186,222  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,205  
FILING DATE: 19-MAR-1991  
APPLICATION NUMBER: GB 9006400.7  
FILING DATE: 22-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Villamizar, JoAnn  
REGISTRATION NUMBER: 30,598  
REFERENCE/DOCKET NUMBER: 4-17994/A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914)785-7121  
TELEFAX: (914)347-5769  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 279 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (recombinant)  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Hybrid gene: MSP signal (base pair)  
INDIVIDUAL ISOLATE: desulfatohirudin (base pair)  
IMMEDIATE SOURCE:  
CLONE: pUCRS/pML310  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..276  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 82..276  
US-08-186-222-3

|                       |                 |                    |           |             |
|-----------------------|-----------------|--------------------|-----------|-------------|
| Query Match           | 78.0%;          | Score 84.2;        | DB 1;     | Length 279; |
| Best Local Similarity | 91.8%;          | Pred. No. 2.7e-20; |           |             |
| Matches 89;           | Conservative 0; | Mismatches 8;      | Indels 0; | Gaps 0;     |

|    |     |  |     |
|----|-----|--|-----|
| QY | 12  | tattcaggtattcttcgggatttctcaagtcgcgcttgcttgtagactcggttc   | 71  |
|    |     |  |     |
|    |     |  |     |
| Db | 278 | TACTCAGGTATTTCTCCGGATTTCTTCAAGTCACCGTCGTTGTGAGACTCGGTTTC | 219 |
|    |     |  |     |
|    |     |  |     |
| QY | 72  | ggagtaccttcgcagtaaacgattgattttttgcg                      | 108 |
|    |     |  |     |
|    |     |  |     |
| Db | 218 | GGGGTACCTTCGCGGGTAAGCGACTGGTGTGTTTTTTCAC                 | 182 |
|    |     |  |     |
|    |     |  |     |

RESULT 8  
5180668-10/c  
; Patent NO. 5180668  
; APPLICANT: CRAUSE, PETER; HABERMANN, PAUL; TRIPIER, DOMINIQUE  
; TITLE OF INVENTION: HIRUDIN DERIVATIVE  
; NUMBER OF SEQUENCES: 10  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/295,422  
; FILING DATE: 10-JAN-1989  
; SEQ ID NO: 10:  
; LENGTH: 212  
5180668-10

```

Query Match          77.0%; Score 83.2; DB 6; Length 212;
Best Local Similarity 87.5%; Pred. No. 5.5e-20;
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy      5  gccctattatgcagggtattcttcgcgggattcttcaaaagtcgcgcgtgtgtgagactg 64
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db      209 GCTCTATTACTGAAGGTATTCTCCTCAGGATCTCTTCGAAGTCGCGCGTGTATGAGACTG 150

Qy      65  cggtttcggagtaccttcgccagtaaacgcattgatttttttcgc 108
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db      149 CGGTTTCGGGTACTTCGCCAGTAACGCACTGGTCTTTTCAC 106

RESULT          9
US-07-982-064-8/c
; Sequence 8, Application US/07982064
; Patent No. 591895
; GENERAL INFORMATION:
; APPLICANT: Schmid, G.; Habermann, P.
; TITLE OF INVENTION: Secretion of Hirudin Derivatives
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Collard, Roe & Galgano, P.C.
; STREET: 1077 No. 591895thern Boulevard
; CITY: Roslyn
; STATE: New York
; COUNTRY: USA
; ZIP: 11576
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.0.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/982,064
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/672,791
; FILING DATE: 21-MAR-1991
; APPLICATION NUMBER: GR 40 09 268.2
; FILING DATE: 22 MAR 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Collard, Allison C.
; REGISTRATION NUMBER: 22,532
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W3
; NAME: Galgano, Thomas M.
; REGISTRATION NUMBER: 27,638
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W3
; NAME: Freedman, Edward R.
; REGISTRATION NUMBER: 26,048
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-365-9802
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 227 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other DNA
; US-07-982-064-8

```

|                       |        |   |       |                |
|-----------------------|--------|---|-------|----------------|
| Query Match           | 77.0%  | Score 83.2;   | DB 2; | Length 227;    |
| Best Local Similarity | 87.5%; | Pred. No. 5.7e-20;  |       |                |
| Matches               | 91;    | Conservative  | 0;    | Mismatches 13; |
|                       |        |   |       | Indels 0;      |
|                       |        |   |       | Gaps 0;        |
| QY                    | 5      | gccctattattgcaggtattcttcgcgggattttctcaagtcgcgcgtgtgtgagactg | 64    |                |
| Db                    | 201    | GCTCTATTACTGAAGTATTCTTCAGGGATCTCTCGAATGCCCGCTGTATGAGACTG    | 142   |                |
| QY                    | 65     | cggtttcggagtagccttcgcagcaagcattgatttttttcgc                 | 108   |                |



Db 141 CGGTTTCGGGGTACCTGCCAGTAAACGCACTGGTTCCTTTTCAC 98  
|||||

## RESULT 10

US-08-757-439-1

Sequence 1, Application US/08757439

Patent No. 5866371

GENERAL INFORMATION:

APPLICANT: BADZONG, Werner

APPLICANT: HABERMANN, Paul

APPLICANT: MOELLER, Joerg

APPLICANT: ARETZ, Werner

TITLE OF INVENTION: PROCESS FOR USING THE YEAST ADH II

TITLE OF INVENTION: PROMOTER SYSTEM FOR THE PRODUCTION OF HETEROLOGOUS

TITLE OF INVENTION: PROTEINS IN HIGH YIELDS

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley &amp; Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/757,439

FILING DATE: 27-NOV-1996

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE 19544233.4

FILING DATE: 28-NOV-1995

ATTORNEY/AGENT INFORMATION:

NAME: SANDERCOCK, Colin G.

REGISTRATION NUMBER: 31,298

REFERENCE/DOCKET NUMBER: 18748/303/HOCE

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 8491 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-757-439-1

## Query Match

Best Local Similarity 77.0%; Score 83.2; DB 2; Length 8491;

Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 5 gccctatttcagggtatttcctccgggatttcctcaaaagtcgcgcgttcgttgagactg 64

Db 6784 GCTCTATTACTGAAGGTATTCTCCAGGGATCTCTTCGAAGTCCGCGTCTATGAGACTG 6843

Qy 65 cggtttcggagtcacctccagtcagtcattgatttttttcgc 108

Db 6844 CGGTTTCGGGGTACCTTCCAGTAAACGCACTGGTTCCTTTTCAC 6887

## RESULT 11

5422249-13/c

Patent No. 5422249

APPLICANT: LIERSCH, MANFRED; RINK, HANS; MARKI, WALTER; GRUTTER,

MARKUS G.; MEYHACK, BERND

TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF THROMBIN

INHIBITORS

NUMBER OF SEQUENCES: 15

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/121,974

FILING DATE: 15-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 700,9978

FILING DATE: 10-MAY-1991

APPLICATION NUMBER: 582,816

FILING DATE: 13-SEP-1990

APPLICATION NUMBER: 211,065

FILING DATE: 20-JUN-1988

APPLICATION NUMBER: 744,453

FILING DATE: 13-JUN-1985

SEQ ID NO:13:

LENGTH: 217

5422249-13

## Query Match

Best Local Similarity 76.5%; Score 82.6; DB 6; Length 217;

Matches 88; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 12 tattgcagggtatttcctccgggatttcctcaaaagtcgcgcgttcgttgagactgcggtttc 71

Db 209 TACTGCAGGTATTCTCCGGGATTCTTCGAAGTCAACGCGTCTGTGAGACTCGCGTTTC 150

Qy 72 ggaqtaccttcgcagtaacgcatgatttttttcgc 108

Db 149 GGGGTACCTTCCCGGTAAACGCACTGGTTTTTTTCAC 113

## RESULT 12

5164304-3/c

Patent No. 5164304

APPLICANT: Johnson, Paul H.; Lazar, Jerome B.; Sohel, Indira

Waleh, Nahid S.

TITLE OF INVENTION: METHOD AND VECTORS FOR STABILIZING HIRUDIN

AND HUMAN LAMININ B1 EXPRESSION

NUMBER OF SEQUENCES: 26

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/347,545

FILING DATE: 04-MAY-1989

SEQ ID NO:3:

LENGTH: 227

5164304-3

## Query Match

Best Local Similarity 75.4%; Score 81.4; DB 6; Length 227;

Matches 88; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 10 attattgcagggtatttcctccgggatttcctcaaaagtcgcgcgttcgttgagactgcggtttc 69

Db 212 ATTATTGCAGGTATTCTCCGGGATCTCTTCGAAGTCCGCGTCTGTGAGACTCGCGTTTC 153

Qy 70 tcgcagtcaccttcgccagtcagtcattgatttttttcgc 108

Db 152 TCGGGGTACCTTCCCGGTAAACGCACTGGTTTTTTTCGC 114

## RESULT 13

5422249-8/c

Patent No. 5422249

APPLICANT: LIERSCH, MANFRED; RINK, HANS; MARKI, WALTER; GRUTTER,

MARKUS G.; MEYHACK, BERND

TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF THROMBIN

INHIBITORS

NUMBER OF SEQUENCES: 15

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/121,974

FILING DATE: 15-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 700,9978

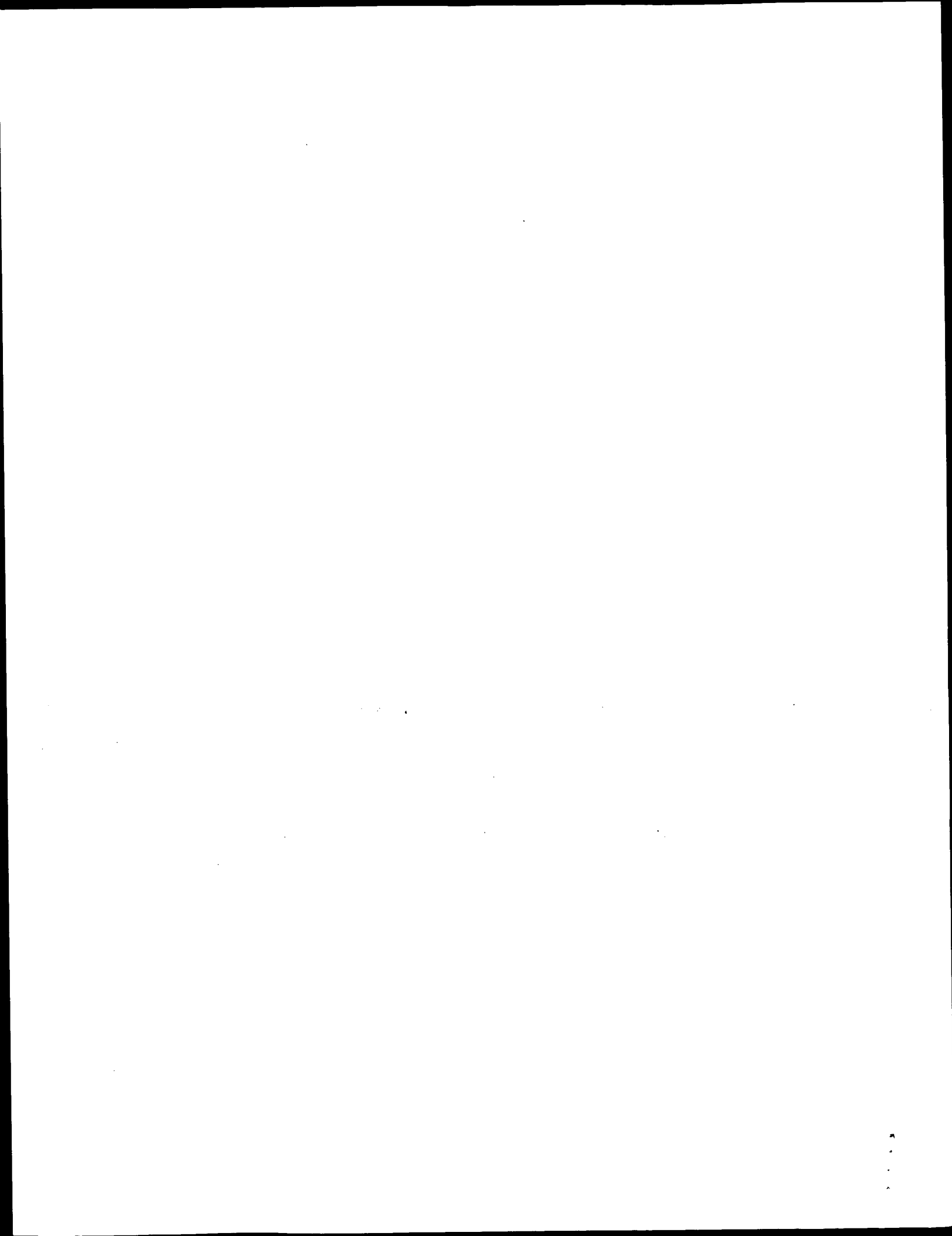


Tue Jul 16 08:24:09 2002

us-10-053-641-4.rni

---

Page 7



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:48:54 ; Search time 6165.88 Seconds  
(without alignments)  
236.409 Million cell updates/sec

Title: US-10-053-641-4

Perfect score: 108

Sequence: 1 ggcgcctattattgcagg.....aacgcattgatttttcgc 108

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpi:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 33    | 30.6        | 621    | 10    | BG671298    |
| 2          | 33    | 30.6        | 644    | 10    | BI107752    |
| 3          | 32.4  | 30.0        | 402    | 12    | TA210G12Q   |
| C 4        | 32.4  | 30.0        | 511    | 12    | TA54D01P    |
| 5          | 29.8  | 27.6        | 678    | 10    | BI798501    |
| C 6        | 29.6  | 27.4        | 978    | 9     | AL554921    |
| C 7        | 29.6  | 27.4        | 1244   | 10    | BG290977    |
| 8          | 29.4  | 27.2        | 387    | 12    | BH214991    |
| C 9        | 29.2  | 27.0        | 352    | 10    | BM218166    |
| 10         | 29.2  | 27.0        | 626    | 9     | AW536772    |
| C 11       | 28.8  | 26.7        | 353    | 9     | AW738920    |
| 12         | 28.8  | 26.7        | 432    | 12    | AZ180488    |
| C 13       | 28.8  | 26.7        | 495    | 10    | BJ174087    |
| C 14       | 28.8  | 26.7        | 551    | 10    | BJ157738    |
| C 15       | 28.8  | 26.7        | 577    | 10    | BJ202386    |
| C 16       | 28.8  | 26.7        | 578    | 10    | BJ202320    |
| C 17       | 28.8  | 26.7        | 579    | 10    | BJ203448    |

|      |      |      |      |    |          |
|------|------|------|------|----|----------|
| C 18 | 28.8 | 26.7 | 580  | 10 | BJ202387 |
| 19   | 28.8 | 26.7 | 601  | 10 | BJ165686 |
| C 20 | 28.6 | 26.5 | 906  | 12 | AZ202441 |
| 21   | 28.6 | 26.5 | 1030 | 12 | CNS034KT |
| C 22 | 28.4 | 26.3 | 200  | 10 | Z28896   |
| 23   | 28.4 | 26.3 | 219  | 9  | AA568870 |
| 24   | 28.4 | 26.3 | 232  | 9  | BB586156 |
| C 25 | 28.4 | 26.3 | 342  | 10 | BF052721 |
| C 26 | 28.4 | 26.3 | 390  | 9  | AA282103 |
| C 27 | 28.4 | 26.3 | 406  | 9  | AA281638 |
| 28   | 28.4 | 26.3 | 435  | 10 | H96697   |
| C 29 | 28.4 | 26.3 | 457  | 9  | AI685214 |
| C 30 | 28.4 | 26.3 | 460  | 10 | BF054388 |
| C 31 | 28.4 | 26.3 | 478  | 9  | AA938721 |
| C 32 | 28.4 | 26.3 | 565  | 9  | AV916968 |
| C 33 | 28.4 | 26.3 | 567  | 10 | BE499455 |
| 34   | 28.4 | 26.3 | 599  | 10 | BI792095 |
| 35   | 28.4 | 26.3 | 699  | 12 | BH547832 |
| 36   | 28.4 | 26.3 | 719  | 9  | AV935444 |
| 37   | 28.4 | 26.3 | 851  | 10 | BE704576 |
| 38   | 28.2 | 26.1 | 354  | 9  | BE039897 |
| 39   | 28.2 | 26.1 | 396  | 9  | AU082915 |
| 40   | 28.2 | 26.1 | 588  | 9  | AU029435 |
| 41   | 28.2 | 26.1 | 609  | 9  | AU097671 |
| 42   | 28.2 | 26.1 | 625  | 10 | BI806759 |
| 43   | 28.2 | 26.1 | 627  | 9  | AU086669 |
| 44   | 28.2 | 26.1 | 669  | 9  | AU076129 |
| 45   | 28.2 | 26.1 | 708  | 9  | AU097672 |

ALIGNMENTS

RESULT 1

BG671298  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

621 bp mRNA linear EST 30-APR-2001  
DRNBOD03 Rat DRG Library Rattus norvegicus cDNA clone DRNBOD03 5', mRNA sequence.  
BG671298  
BG671298.1 GI:13893397  
EST.  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 621)  
Xiao,H.S., Han,Z.G., Zhang,F.X., Huang,Q.H., Lu,Y.J., Bao,L., Fu,G., Guo,C., Yan,Q., Jin,S.X., Zhu,Z.D., Xu,X.R., Li,N.G., Chen,Z. and Zhang,X.  
Distinct gene expression profiles of rat dorsal root ganglion induced by peripheral nerve axotomy  
Unpublished (2001)  
Contact: Zhang Xu  
Laboratory of Sensory System  
Institute of Neuroscience  
320 Yue Yang Road, Shanghai 200031, P.R.China  
Tel: 86-21-64748700-121  
Fax: 86-21-64713446  
Email: xu.zhang@ion.ac.cn  
This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)  
PCR Primers  
FORWARD: T3  
BACKWARD: T7  
Seq primer: T3  
POLYA=No.  
Location/Qualifiers  
1..621  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"

genomic survey sequence.  
 AL457164  
 ACCESSION  
 AL457164.1  
 VERSION  
 GI:11857555

```

Seq primer: M13 forward primer.
Location/Qualifiers
    i. 678
    /organism="Oryza sativa"
    /db_xref="taxon:4530"
    /clone="H109E08"
    /clone_lib="Endosperm library from Oryza sativa (10 days after anthesis)"
    /tissue_type="Endosperm"
    /dev_stage="10 days after anthesis"
    /note="vector: pSport2"

BASE COUNT      159 a      180 c      169 g      170 t
ORIGIN

Query Match          27.6%; Score 29.8; DB 10; Length 678;
Best Local Similarity 56.7%; Pred. No. 44;
Matches 55; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY   12  tattcaggtattcttcgggattttcttcaagtcgcgcgtcggtgtgagactgcggtttc 71
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    435 TAGTGTAGTAGTGTTGGTGATTTCTTCGTCCTCGTGAGGAGTCGCATGCCAGC 494

QY   72  ggagtacctccaggtaacgcattgatTTTTTctgc 108
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    495 GGGTGTAATTGCCCTGTGCATCTCTGATTTCTTTCTC 531

RESULT           6
AL554921/c
LOCUS            AL554921
DEFINITION       AL554921 LTI_NFL006_P12 Homo sapiens cDNA clone CS0DI087P21 5' linear EST 16-FEB-2001
```

```

Human1.
SOURCE ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 978)
AUTHORS Li, W. B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqraf@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
    source
        1..978
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CS0D1087YP21"
            /clone_lib="LTI_NFL006_PL2"
            /tissue_type="placenta"
            /notes="Vector: pCMVSPORT 6; Site1: NotI; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            was priched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed by
            Life Technologies. Contact : Feng Liang Life Technologies,
            a division of Invitrogen 9800 Medical Center Drive
            Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
            Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com"
            320 a 223 c 235 g 190 t 10 others
BASE COUNT
ORIGIN
Query Match 27.4%; Score 29.6; DB 9; Length 978;
Best Local Similarity 62.5%; Pred. No. 54;
Matches 45; Conservative 1; Mismatches 26; Indels 0; Gaps 0;

```

```

Qy 5 gccctatttcgaggtatttccgggatttcttccaaagtcgccgctgtgtgagactg 64
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 206 GGCCTCTGTTGGTATTTTCTCGAGTCTCTGCCAAGTGGVGTGAGATACAGACTG 147
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 65 cgggttcggagt 76
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146 CTGAGTTATAGT 135

RESULT 7
BG290977/c
LOCUS BG290977 1244 bp mRNA linear EST 21-FEB-2001
DEFINITION 602386963F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516023 5',
mRNA sequence.
ACCESSION BG290977
VERSION BG290977.1 GI:13048483
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
CDNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10406 row: m column: 16
High quality sequence stop: 653.
Location/Qualifiers
1. .1244
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4516023"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: bladder; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 406 a 312 c 247 g 279 t
ORIGIN

Query Match 27.4%; Score 29.6; DB 10; Length 1244;
Best Local Similarity 61.8%; Pred. No. 56;
Matches 47; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 17 cagggtatttccgggatttccaaagtcgccgctgtgtgagactgagggttcggagt 76
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 921 CCGGGGTCCTGTTGTTGTTTCAATGTCGGCGTGTGAGAAATGCAGATGTAGGACC 862
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 77 accttcgcagtaacg 92
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 861 CATTTTGCACGAATG 846
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
BH214991
LOCUS BH214991 387 bp DNA linear GSS 08-NOV-2001
DEFINITION 1006012F09.2EL_x1 1006 - RescueMu Grid G Zea mays genomic, DNA
sequence.
ACCESSION BH214991
VERSION BH214991.1 GI:16805681

```

```

KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
JOURNAL Clade; Panicoideae; Andropogoneae; Zea.
COMMENT 1 (bases 1 to 387)
Walbot,V.
Maize genomic sequences found using engineered RescueMu transposon
Unpublished (2001)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1006012 row: 38
Class: transposon-tagged.
Location/Qualifiers
1. .387
/organism="Zea mays"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/clone_lib="1006 - RescueMu Grid G"
/dev_stage="adult"
/tissue_type="leaf"
/lab_host="DH10B"
/notes="Organ: leaf; Vector: RescueMu (engineered from
pBluescript backbone); Site_1: BamHI; Site_2: BglII;
RescueMu is a 4.9 kb, modified maize Mu transposon
designed to allow plasmid rescue from total genomic DNA.
Mu elements insert preferentially into transcription
units. For more information on RescueMu, go to the web
site 'www.zmdb.iastate.edu' and follow the links for
'RescueMu.' Grid G was grown at Stanford in 2000. DNA was
extracted from leaf punches, double digested using BamHI
and BglII, and ligated to form circular plasmids. DH10B
cells were transformed and then screened on LB plates with
ampicillin."
BASE COUNT 101 a 98 c 94 g 93 t 1 others
ORIGIN

Query Match 27.2%; Score 29.4; DB 12; Length 387;
Best Local Similarity 56.8%; Pred. No. 54;
Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 4 cggcctatttcgaggtatttccgggatttcttccaaagtcgccgctgtgtgagact 63
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36 CCCCACATTGATCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 95
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 64 gcgggttcggagtagcttcgagtagcttcgagtagcttcgagtagcttcgagtagct 98
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96 GAGGCCCGGAGACGCTTCAGTAGGATCTGAGGGA 130
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
BM218166/c
LOCUS BM218166 352 bp mRNA linear EST 31-JAN-2002
DEFINITION C0907B09-3 NIA Mouse 12.5-dpc Male Genital Ridge/Mesonephros cDNA
Library (Long) Mus musculus cDNA clone C0907B09 3', mRNA sequence.
ACCESSION BM218166
VERSION BM218166.1 GI:17777596
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 352)
Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Pantano,S., Luo,A.

```



TITLE and Ko,M.S.H.  
 JOURNAL Systematic Analyses of NIA Mouse 12.5-dpc Male Genital  
 COMMENT Ridge/Mesonephros cDNA Library (Long)  
 Unpublished (2001)  
 Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@lgsun.grc.nia.nih.gov  
 Plate: C0907 row: B column: 09  
 Seq primer: -21M13 Forward  
 High quality sequence stop: 352  
 POLYA=Yes.

## FEATURES

source  
 1. 352  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="niaEST:C0907B09-3"  
 /db\_xref="taxon:10090"  
 /clone="C0907B09"  
 /clone\_lib="NIA Mouse 12.5-dpc Male Genital  
 Ridge/Mesonephros cDNA Library (Long)"  
 /sex="Male"  
 /tissue\_type="Male genital ridge/mesonephros"  
 /dev\_stage="12.5-dpc"  
 /lab\_host="DH10B"  
 /note="Vector: pSPORT1 (Invitrogen); Site\_1: SalI; Site\_2:  
 NotI; Mouse cDNA project by the Laboratory of Genetics,  
 National Institute on Aging (NIA), Intramural Research  
 Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is  
 a long-transcript enriched cDNA library (Ref. Genome Res.  
 11: 1553-1558 (2001). [PMID: 11544199]). Double-stranded  
 cDNAs were synthesized with an Oligo(dT) primer  
 [Invitrogen]:  
 5'-pGACTAGTCTAGTCGCGAGCGCGCCCTTTTTTTTTTTT-3' from  
 1.8 ug of total RNA, treated with T4 DNA polymerase, and  
 purified by ethanol-precipitation. The cDNAs were ligated  
 to Lone-linker LL-Sal4, purified by phenol/chloroform, and  
 separated from free linkers by Centricon 100. Then, the  
 cDNAs were amplified by long-range high fidelity PCR using  
 Ex Taq polymerase (Takara) with a primer Sal4-S. The  
 products were purified by phenol/chloroform and Centricon  
 100. The cDNAs were digested with SalI and NotI enzymes  
 and cloned into SalI/NotI site of pSPORT1 plasmid vector.  
 The DH10B E. coli host was transformed with the ligation  
 mixture by the standard chemical method. The average  
 insert size is about 2.4 kb. The library was constructed  
 by Fulan Piao (NIA)."  
 BASE COUNT 136 a 65 c 50 g 101 t  
 ORIGIN

Query Match 27.0%; Score 29.2; DB 10; Length 352;  
 Best Local Similarity 56.1%; Pred. No. 62;  
 Matches 55; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
 QY 8 ctattatgcaggtattcttcgggattcttcaaacgcgcgtgtgtgagactcgg 67  
 Db 144 CTAATAGTCTAGTAATTGGTGGCGCTTGAAGTAAAGCTTCTTGTGACTGTGATTTCCT 85  
 QY 58 ttccggagtaaccttcgccagtaacgcattgatttttt 105  
 Db 84 TTTGGTGTATTATTGCTAAGTGAACCTTGTAAATTT 47

RESULT 10  
 AW536772  
 LOCUS 626 bp mRNA linear EST 31-AUG-2000  
 DEFINITION AW536772 NIA Mouse E7.5 Embryonic Portion cDNA Library Mus  
 musculus  
 AW536772  
 ACCESSION cDNA clone G0108E12 3', mRNA sequence.  
 VERSION  
 AW536772.1 GI:7179189  
 KEYWORDS EST.

## SOURCE

ORGANISM house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 626)  
 REFERENCE Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac  
 AUTHORS M.J., Pantano,S., Sano,Y., Piao,Y., Nagara,R., Doi,H., Wood,W.H.  
 III, Becker,K.G. and Ko,M.S.H.  
 TITLE Genome-wide expression profiling of mid-gestation placenta and  
 embryo using a 15,000 mouse developmental cDNA microarray  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)  
 MEDLINE 20381348  
 COMMENT Contact: George J. Kargul  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@lgsun.grc.nia.nih.gov  
 Plate: G0108 row: E column: 12  
 Seq primer: -21M13 Forward  
 High quality sequence stop: 626  
 POLYA=Yes.

## FEATURES

source  
 1. 626  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="niaEST:G0108E12-3"  
 /db\_xref="taxon:10090"  
 /clone="G0108E12"  
 /clone\_lib="NIA Mouse E7.5 Embryonic Portion cDNA Library"  
 /sex="unknown"  
 /dev\_stage="7.5dpc Embryo"  
 /lab\_host="DH10B"  
 /note="Vector: pSPORT1 (Gibco/BRL Life Technology);  
 Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from  
 6 Embryo. The double-stranded cDNA was synthesized by  
 Gibco's kit with an Oligo(dT) primer [NotI primer-adaptor  
 from GibcoBRL]  
 [5'-pGACTAGTCTAGTCGCGAGCGCGCCCTTTTTTTTTTTT-3']  
 from 0.5ug of mRNA. The double-stranded cDNAs were  
 treated with T4 DNA polymerase and purified by  
 ethanol-precipitation. The cDNAs were ligated to  
 Lone-linker LL-Sal3 (include SalI sequence). The cDNAs  
 were purified by phenol/chloroform and separated from  
 free linkers by Centricon 100. Then, cDNAs were amplified  
 by long-range high fidelity PCR using Takara's Ex Taq  
 polymerase. Then, the cDNAs were purified by  
 phenol/chloroform and by Centricon 100. The cDNAs were  
 digested with SalI and NotI enzymes. Then, the cDNAs were  
 size selected by Gibco's Size Fractionation Column. The  
 cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid  
 vector. The DH10B E. coli host was transformed with the  
 ligation mixture by chemical method. The library was  
 constructed by Xiaohong Wang and Minoru S. H. Ko."  
 BASE COUNT 220 a 132 c 123 g 151 t  
 ORIGIN

Query Match 27.0%; Score 29.2; DB 9; Length 626;  
 Best Local Similarity 62.2%; Pred. No. 67;  
 Matches 46; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
 QY 15 tgcaggtattcttcgggattcttcaaacgcgcgtgtgtgagactcgggttcgga 74  
 Db 510 TGATGGCATTTCTCCAGGCTCTCTTCAAGCCAGCCCTACCTGATGAGACCTTGGAAATGGGA 569  
 QY 75 gtaccttcgccagt 88  
 Db 570 AAACTGCATCAAT 583

RESULT 11  
 AW738920/c  
 LOCUS AW738920 353 bp mRNA linear EST 25-APR-2000





Tel: 81-559-81-6856  
Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp

A backbone of the vector is pBluescript II, that was in vivo excised from a modified LPS phage vector (Mo bi Tec, Germany). XhoI digested-5' end of cDNA is ligated to SalI site of the vector, and the BamHI digested-3' end including poly-A tail is ligated to BamHI site of the vector. cDNA insert could be amplified with conventional T7 and T3 primers. This normalized full-length cDNA library was generated basically according to the method described in Genome Research 10, 1617-1630 (2000), Carninci, P. et al. Protonemata were blended by the POLYTRON, and then cultivated on the BCD medium containing 100 µM NAA (naphthalene acetic acid) for 8 to 11 days under the continuous light.

# FEATURES

source

1. .577  
/organism="Physcomitrella patens subsp. patens"  
/db\_xref="taxon:145481"  
/clone="pph45122"  
/clone\_lib="normalized full length cDNA library,  
chloronemata, caulonemata and rhizoid-like protonemata"  
/tissue\_type="mixture of chloronemata, caulonemata and  
rhizoid-like protonemata"

BASE COUNT 121 a 118 c 157 g 181 t

ORIGIN

Query Match 26.7%; Score 28.8; DB 10; Length 577;  
Best Local Similarity 60.0%; Pred. No. 88;  
Matches 48; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
Qy 26 ttccgggattcttcaaaagtcgcgtgtgtgagactgcgggttcggaggtacattcgcc 85  
Db 291 TACTGGTTCTTCTTAAATCGTCGTCTGGACACTGCCCTCTCTGTCAGGTTCTTC 232  
Qy 86 agtaacgcatgtattttt 105  
Db 231 TGGAACTTAGCGCTCTTCTT 212

Search completed: July 15, 2002, 21:49:01  
Job time: 1892 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:28:44 ; Search time 2368.24 Seconds  
(without alignments)  
291.599 Million cell updates/sec

Title: US-10-053-641-5  
Perfect score: 33  
Sequence: 1 tcggatcccttattggtttacactgactgc 33

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pi:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query | Score | Match | Length | DB | ID | Description |
|--------|-------|-------|-------|--------|----|----|-------------|
|--------|-------|-------|-------|--------|----|----|-------------|

|      |      |      |        |    |           |                    |
|------|------|------|--------|----|-----------|--------------------|
| 1    | 25.2 | 76.4 | 226    | 12 | SYNHIR    | M26762 Synthetic h |
| 2    | 23.2 | 70.3 | 150214 | 9  | AC011291  | AC011291 Homo sapi |
| c 3  | 23.2 | 70.3 | 188732 | 2  | AC024050  | AC024050 Homo sapi |
| 4    | 23.2 | 70.3 | 188237 | 2  | AP000906  | AP000906 Homo sapi |
| 5    | 23.2 | 70.3 | 196695 | 2  | AC078936  | AC078936 Homo sapi |
| c 6  | 23.2 | 70.3 | 269711 | 2  | AP000408  | AP000408 Homo sapi |
| 7    | 22.4 | 67.9 | 33     | 6  | A00633    | A00633 Nucleotide  |
| 8    | 22.4 | 67.9 | 220    | 12 | SYNHIRMED | M14964 Synthetic h |
| 9    | 22.4 | 67.9 | 224    | 6  | A00630    | A00630 Artificial  |
| c 10 | 22.4 | 67.9 | 224    | 6  | A00631    | A00631 Artificial  |
| 11   | 22.4 | 67.9 | 224    | 6  | A04277    | A04277 Artificial  |
| c 12 | 22.4 | 67.9 | 224    | 6  | A04278    | A04278 Artificial  |
| 13   | 22.4 | 67.9 | 224    | 6  | A07345    | A07345 Synthetic D |
| c 14 | 22.4 | 67.9 | 224    | 6  | A07346    | A07346 Synthetic D |
| c 15 | 22   | 66.7 | 34     | 6  | A00634    | A00634 Nucleotide  |
| 16   | 22   | 66.7 | 93     | 6  | I63540    | I63540 Sequence 3  |
| 17   | 21.8 | 66.1 | 80291  | 3  | AC096493  | AC096493 Rattus no |
| 18   | 21.4 | 64.8 | 35875  | 3  | CE2C15    | Z93396 Caenorhabdi |
| c 19 | 21.4 | 64.8 | 155550 | 2  | AC020227  | AC020227 Drosophil |
| c 20 | 21.4 | 64.8 | 162387 | 3  | AC008233  | AC008233 Drosophil |
| 21   | 21.4 | 64.8 | 171952 | 10 | AC002324  | AC002324 Mus muscu |
| 22   | 21.4 | 64.8 | 192550 | 2  | AC025584  | AC025584 Mus muscu |
| c 23 | 21.4 | 64.8 | 217238 | 2  | AL591131  | AL591131 Mus muscu |
| 24   | 21.4 | 64.8 | 232750 | 2  | AL591177  | AL591177 Mus muscu |
| c 25 | 21.4 | 64.8 | 241429 | 3  | AE003608  | AE003608 Drosophil |
| c 26 | 21.2 | 64.2 | 84478  | 9  | AL133462  | AL133462 Human DNA |
| c 27 | 21   | 63.6 | 106    | 6  | I41186    | I41186 Sequence 6  |
| 28   | 21   | 63.6 | 129    | 6  | I41183    | I41183 Sequence 3  |
| 29   | 21   | 63.6 | 235    | 6  | I41187    | I41187 Sequence 7  |
| c 30 | 21   | 63.6 | 235    | 6  | I41188    | I41188 Sequence 8  |
| 31   | 21   | 63.6 | 726    | 6  | I41189    | I41189 Sequence 9  |
| c 32 | 21   | 63.6 | 726    | 6  | I41190    | I41190 Sequence 10 |
| c 33 | 21   | 63.6 | 4809   | 6  | AX200939  | AX200939 Sequence  |
| c 34 | 21   | 63.6 | 4809   | 6  | AX267595  | AX267595 Sequence  |
| c 35 | 21   | 63.6 | 72273  | 2  | AC090859  | AC090859 Homo sapi |
| c 36 | 21   | 63.6 | 143420 | 9  | AL161422  | AL161422 Human DNA |
| 37   | 20.8 | 63.0 | 58     | 6  | A34231    | A34231 Synthetic h |
| 38   | 20.8 | 63.0 | 109    | 6  | A34234    | A34234 Synthetic h |
| c 39 | 20.8 | 63.0 | 109    | 6  | A34235    | A34235 Synthetic h |
| 40   | 20.8 | 63.0 | 217    | 6  | A03695    | A03695 Nucleotide  |
| c 41 | 20.8 | 63.0 | 217    | 6  | A03696    | A03696 Nucleotide  |
| 42   | 20.8 | 63.0 | 217    | 6  | A34238    | A34238 Synthetic d |
| c 43 | 20.8 | 63.0 | 217    | 6  | A34239    | A34239 Synthetic d |
| 44   | 20.8 | 63.0 | 217    | 6  | E00657    | E00657 DNA encodin |
| 45   | 20.8 | 63.0 | 168111 | 9  | HS525L6   | AL023807 Human DNA |

ALIGNMENTS

|            |        |   |           |        |     |        |                 |
|------------|--------|---|-----------|--------|-----|--------|-----------------|
| RESULT     | 1      | SYNHIR  | SYNHIR    | 226 bp | DNA | linear | SYN 27-APR-1993 |
| LOCUS      | SYNHIR | Synthetic hirudin gene, complete cds.                               |           |        |     |        |                 |
| DEFINITION |        | M26762  |           |        |     |        |                 |
| ACCESSION  |        | M26762.1  | GI:208478 |        |     |        |                 |
| VERSION    |        | hirudin; proteinase inhibitor; thrombin inhibitor.                  |           |        |     |        |                 |
| KEYWORDS   |        | Synthetic DNA.  |           |        |     |        |                 |
| SOURCE     |        | artificial construct  |           |        |     |        |                 |
| ORGANISM   |        | Bergmann, C., Dodt, J., Koehler, S., Fink, E. and Gassen, H. G.     |           |        |     |        |                 |
| REFERENCE  |        | 1 (bases 1 to 226)  |           |        |     |        |                 |
| AUTHORS    |        | Chemical synthesis and expression of a gene coding for hirudin, the |           |        |     |        |                 |
| TITLE      |        | thrombin-specific inhibitor from the leech Hirudo medicinalis       |           |        |     |        |                 |
| JOURNAL    |        | Biol. Chem. Hoppe-Seyler 367, 731-740 (1986)                        |           |        |     |        |                 |
| MEDLINE    |        | 87026239  |           |        |     |        |                 |
| FEATURES   |        | Location/Qualifiers   |           |        |     |        |                 |
| source     |        | 1..226  |           |        |     |        |                 |
| CDS        |        | /organism="synthetic construct"                                     |           |        |     |        |                 |
|            |        | /db_xref="taxon:32630"  |           |        |     |        |                 |
|            |        | 22..222   |           |        |     |        |                 |
|            |        | /note="hirudin"   |           |        |     |        |                 |
|            |        | /codon_start=1  |           |        |     |        |                 |

```

/transl_table=11
/protein_id="AAA72772.1"
/db_xref="GI:208479"
/translation="MVVYDCTESGNLCLCEGSGVCGGNKCLGSDGCKNQCVTGE
GTPKQSHNDGDFEIPPEYIQ"
BASE COUNT      63 a   56 c   61 g   46 t
ORIGIN

```

```

Query Match      76.4%; Score 25.2; DB 12; Length 226;
Best Local Similarity 90.0%; Pred. No. 1.4;
Matches 27; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 4 ggatccttattgtgtttacactgactgc 33
| | | | | | | | | | | | | | | | | |
Db 13 GTAAGCTTTATGGTTGTTTACACTGACTGC 42

```

```

RESULT 2
AC011291 AC011291 150214 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-67G7 from 2, complete sequence.
DEFINITION
AC011291
AC011291
AC011291.8 GI:14589665
VERSION
HTG.
SOURCE
human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 150214)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792

```

```

REFERENCE 2 (bases 1 to 150214)
AUTHORS Mulvaney,E., Boyer,E. and Kozlowski,A.
TITLE The sequence of Homo sapiens BAC clone RP11-67G7
JOURNAL Unpublished (2001)

```

```

REFERENCE 3 (bases 1 to 150214)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

REFERENCE 4 (bases 1 to 150214)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

REFERENCE 5 (bases 1 to 150214)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

```

```

COMMENT On Jul 3, 2001 this sequence version replaced gi:14018119.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
-----
Summary Statistics
-----
Center project name: H_NH0067G07

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenio,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-693G13, 2000 bp overlap; the clone sequenced to the right is RP11-434M17, 2000 bp overlap. Actual end of this clone is at base position 14557 of RP11-434M17.

The sequence of RP11-67G7 from 94121 to 94182 is from a PCR product of clone DNA.

| FEATURES      | Location/Qualifiers   |
|---------------|---|
| source        | 1..150214<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/chromosome="2"<br>/map="2" |
| repeat_region | /clone="RP11-67G7"<br>/clone.lib="RPCI-11"<br>831..1416<br>/rpt_family="L2"                   |
| misc_feature  | 1832..2326<br>/note="match to EST AA885327 (NID:g2994404) al60g03.s1"                         |
| misc_feature  | 1833..2321<br>/note="match to EST BG196474 (NID:g13718161)"                                   |
| misc_feature  | 1837..2120<br>/note="match to EST H96169 (NID:g1109311) yt98b10.s1"                           |
| misc_feature  | 1849..2114<br>/note="match to EST BG206827 (NID:g13728514)"                                   |
| misc_feature  | 2161..2386<br>/note="similar to Homo sapiens EST AA757626 (NID:g2805489) 2942a02.s1"          |
| repeat_region | 2325..2345<br>/rpt_family="(TG)n"   |
| misc_feature  | 3080..3662<br>/note="match to EST H23520 (NID:g8922215) ym53d01.r1"                           |
| misc_feature  | 3167..3435<br>/note="match to EST R50919 (NID:g812821) yg70h03.r1"                            |
| repeat_region | 3399..3515<br>/rpt_family="CT-rich"   |
| repeat_region | 3537..4947<br>/rpt_family="L1"  |
| repeat_region | 4948..5121<br>/rpt_family="L1"  |
| repeat_region | 5873..6004<br>/rpt_family="MIR"   |
| repeat_region | 6006..6096<br>/rpt_family="CR1"   |
| repeat_region | 6100..6137<br>/rpt_family="(TTTTTA)n"   |
| repeat_region | 6109..6432<br>/rpt_family="Alu"   |
| repeat_region | 7267..7327<br>/rpt_family="L1"  |
| repeat_region | 7475..7506<br>/rpt_family="AT-rich"   |

```

repeat_region 8933. .8984
/rpt_family="T)n"
repeat_region 11892. .11965
/rpt_family="MIR"
repeat_region 12710. .12914
/rpt_family="L2"
repeat_region 12950. .13072
/rpt_family="MIR"
repeat_region 13265. .13302
/rpt_family="TATG)n"
misc_feature 13599. .14264
/note="match to EST AL557921 (NID:gi12901993)"
misc_feature 13600. .13852
/note="match to EST BG184202 (NID:gi13705889)"
misc_feature 13656. .14143
/note="match to EST H94889 (NID:gl102522) yu57g08.s1"
misc_feature 13863. .14766
/note="match to EST AL579862 (NID:gi12945319)"
repeat_region 14186. .14213
/rpt_family="AT-rich"
repeat_region 14211. .14312
/rpt_family="MIR"
misc_feature 14782. .14785
/note="match to EST AL579862 (NID:gi12945319)"
repeat_region 14806. .14827
/rpt_family="AT-rich"
misc_feature 15159. .15479
/note="match to EST AA774008 (NID:g2825897) ab67f02.r1"
misc_feature 15302. .15776
/note="match to EST A1277366 (NID:g3899634) qm54b12.x1"
repeat_region 15861. .15966
/rpt_family="(TA)n"
repeat_region 16083. .16277
/rpt_family="Alu"
repeat_region 16368. .16473
/rpt_family="Alu"
repeat_region 16445. .16464
/rpt_family="(CAAAA)n"
repeat_region 16598. .17048
/rpt_family="(TG)n"
repeat_region 17257. .17330
/rpt_family="(TA)n"
repeat_region 17570. .17724
/rpt_family="MIR"
repeat_region 18450. .18505
/rpt_family="L2"
repeat_region 18852. .19013
/rpt_family="MER1_type"
repeat_region 19224. .19351
/rpt_family="MER1_type"
repeat_region 19311. .19418
/rpt_family="GA-rich"
repeat_region 19443. .19510
/rpt_family="MER1_type"
repeat_region 19851. .20074
/rpt_family="MIR"
repeat_region 20084. .20118
/rpt_family="Mariner"
repeat_region 20201. .20258
/rpt_family="MIR"
repeat_region 20305. .20409
/rpt_family="L2"
misc_feature 20666. .20679
/note="match to EST BG214571 (NID:gi13740592)"
repeat_region 20666. .20959
/rpt_family="Alu"
repeat_region 21013. .21049
/rpt_family="(CA)n"
repeat_region 21508. .21544
/rpt_family="(TG)n"
repeat_region 23468. .23634
/rpt_family="MIR"
repeat_region 24535. .24643

```

```

Query Match 70.3%; Score 23.2; DB 9; Length 150214;
Best Local Similarity 89.3%; Pred. No. 7;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ggatcctttatggtttttactgact 31
||||| | |||||||||
Db 6647 GGATCCTGTTAGTTGTTTACTGACT 6674

RESULT 3
AC024050/c
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-569A20, WORKING DRAFT
AC024050
AC024050.6 GI:7637374
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 168732)
Waterston,R.H.
Unpublished
The sequence of Homo sapiens clone
2 (bases 1 to 168732)
Waterston,R.H.
Direct Submission
Submitted (20-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Apr 22, 2000 this sequence version replaced gi:7579880.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0569A20
----- Summary Statistics -----
Sequencing vector: pLM3; 77%
Chemistry: Dye-primer ET; 77% of reads
Assembly: Dye-terminator Big Dye; version 0.990319
Consensus quality: 166423 bases at least Q40
Consensus quality: 166920 bases at least Q30
Consensus quality: 167253 bases at least Q20
Insert size: 166000; agarose-fp
Quality size: 168332; sum-of-contigs
Quality coverage: 7.51 in Q20 bases; agarose-fp
Quality coverage: 7.44 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1680: contig of 1680 bp in length
* 1681 1780: gap of unknown length
* 1781 9168: contig of 7388 bp in length
* 9169 9269: gap of unknown length
* 9269 26141: contig of 16873 bp in length
* 26142 26241: gap of unknown length
* 26242 87019: contig of 60778 bp in length
* 87020 87119: gap of unknown length
* 87120 168732: contig of 81613 bp in length.
* Location/Qualifiers

```

FEATURES





100504 105342: contig of 4839 bp in length  
105343 105442: gap of 100 bp  
105443 110335: contig of 4893 bp in length  
110336 110435: gap of 100 bp  
110436 115335: contig of 4900 bp in length  
115336 115435: gap of 100 bp  
115436 119336: contig of 4101 bp in length  
119337 119636: gap of 100 bp  
119637 124058: contig of 4422 bp in length  
124059 124158: gap of 100 bp  
124159 128684: contig of 4526 bp in length  
128685 128784: gap of 100 bp  
128785 132388: contig of 3604 bp in length  
132389 132488: gap of 100 bp  
132489 136139: contig of 3651 bp in length  
136140 136239: gap of 100 bp  
136240 139193: contig of 2954 bp in length  
139194 139293: gap of 100 bp  
139294 142207: contig of 2914 bp in length  
142208 142307: gap of 100 bp  
142308 146335: contig of 4028 bp in length  
146336 146435: gap of 100 bp  
146436 149853: contig of 3418 bp in length  
149854 149953: gap of 100 bp  
149954 152741: contig of 2788 bp in length  
152742 152841: gap of 100 bp  
152842 154497: contig of 1656 bp in length  
154498 154597: gap of 100 bp  
154598 156939: contig of 2342 bp in length  
156940 157039: gap of 100 bp  
157040 157559: contig of 520 bp in length  
157560 157659: gap of 100 bp  
157660 159511: contig of 1852 bp in length  
159512 159611: gap of 100 bp  
159612 162263: contig of 2652 bp in length  
162264 162363: gap of 100 bp  
162364 164148: contig of 1785 bp in length  
164149 164248: gap of 100 bp  
164249 166170: contig of 1922 bp in length  
166171 166270: gap of 100 bp  
166271 168664: contig of 2394 bp in length  
168665 168764: gap of 100 bp  
168765 170677: contig of 1913 bp in length  
170678 170777: gap of 100 bp  
170778 173092: contig of 2315 bp in length  
173093 173192: gap of 100 bp  
173193 175018: contig of 1826 bp in length  
175019 175118: gap of 100 bp  
175119 176699: contig of 1581 bp in length  
176700 176799: gap of 100 bp  
176800 178681: contig of 1882 bp in length  
178682 178781: gap of 100 bp  
178782 180124: contig of 1343 bp in length  
180125 180224: gap of 100 bp  
180225 182291: contig of 2067 bp in length  
182292 182391: gap of 100 bp  
182392 183683: contig of 1292 bp in length  
183684 183783: gap of 100 bp  
183784 185242: contig of 1459 bp in length  
185243 185342: gap of 100 bp  
185343 186827: contig of 1485 bp in length  
186828 186927: gap of 100 bp  
186928 188237: contig of 1310 bp in length.

FEATURES  
Source

1. .188237  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11q22"  
/clone="RP11-70717"  
1. .14984  
misc\_feature  
/note="assembly\_fragment"  
15085. .25511  
misc\_feature

/note="assembly\_fragment"

Query Match 70.3%; Score 23.2; DB 2; Length 188237;  
Best Local Similarity 89.3%; Pred. No. 6.9;  
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 5 gatcctttatgattgttttccctgactg 32  
DB 109550 GATCCTTTATGATTGTTTCCCTGACTG 109577

RESULT 5

AC078936 196695 bp DNA linear HTG 11-AUG-2000  
Homo sapiens chromosome 2 clone RP11-69G4 map 2, WORKING DRAFT  
SEQUENCE, 13 unordered pieces.  
AC078936  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
AC078936.1 GI:9795655  
human.  
ORGANISM  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 196695)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 2, clone RP11-69G4  
Unpublished  
2 (bases 1 to 196695)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Callymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galgan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,M., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehotzky,J.,  
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Melgrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenda,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rohtman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (11-AUG-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submission@genome.wi.mit.edu  
----- Project Information  
Center project name: L9969  
Center clone name: 69.G.4  
----- Summary Statistics  
Sequencing vector: M13; W77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 191343 bases at least Q40  
Consensus quality: 194310 bases at least Q30  
Consensus quality: 195161 bases at least Q20  
Insert size: 185000; agarose-fp  
Insert size: 195495; sum-of-ctnigs  
Quality coverage: 5.4 in Q20 bases; agarose-fp

TITLE  
JOURNAL  
COMMENT

Tue Jul 16 08:24:10 2002

Quality coverage: 5.2 in Q20 bases; sum-of-contigs

\*\*\*\*\*  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 13 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 14276: contig of 14276 bp in length  
 \* 14277 14376: gap of 100 bp  
 \* 14377 15570: contig of 1194 bp in length  
 \* 15571 15670: gap of 100 bp  
 \* 15671 17173: contig of 1503 bp in length  
 \* 17174 17273: gap of 100 bp  
 \* 17274 20185: contig of 2912 bp in length  
 \* 20186 20285: gap of 100 bp  
 \* 20286 277045: contig of 56760 bp in length  
 \* 77046 77145: gap of 100 bp  
 \* 77146 80660: contig of 3515 bp in length  
 \* 80661 80760: gap of 100 bp  
 \* 80761 84452: contig of 3692 bp in length  
 \* 84453 84552: gap of 100 bp  
 \* 84553 88612: contig of 4060 bp in length  
 \* 88613 88712: gap of 100 bp  
 \* 88713 93471: contig of 4759 bp in length  
 \* 93472 93571: gap of 100 bp  
 \* 93572 103205: contig of 9634 bp in length  
 \* 103206 103305: gap of 100 bp  
 \* 103306 136665: contig of 3360 bp in length  
 \* 136666 136765: gap of 100 bp  
 \* 136766 193339: contig of 56574 bp in length  
 \* 193340 193439: gap of 100 bp  
 \* 193440 196695: contig of 3256 bp in length.

## FEATURES

source  
 1..196695  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="RP11-69G4"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1..14276  
 /note="assembly\_fragment"  
 clone\_end:sp6  
 vector\_side:left  
 14377..15570  
 /note="assembly\_fragment"  
 15671..17173  
 /note="assembly\_fragment"  
 17274..20185  
 /note="assembly\_fragment"  
 20286..77045  
 /note="assembly\_fragment"  
 77146..80660  
 /note="assembly\_fragment"  
 80761..84452  
 /note="assembly\_fragment"  
 84553..88612  
 /note="assembly\_fragment"  
 88713..93471  
 /note="assembly\_fragment"  
 93572..103205  
 /note="assembly\_fragment"  
 103306..136665  
 /note="assembly\_fragment"  
 136766..193339  
 /note="assembly\_fragment"  
 193440..196695  
 /note="assembly\_fragment"  
 clone\_end:t7  
 vector\_side:right

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

BASE COUNT 61880 a 36045 c 35392 g 62178 t 1200 others  
 ORIGIN

Query Match 70.3%; Score 23.2; DB 2; Length 196695;  
 Best Local Similarity 89.3%; Pred. No. 6.9;  
 Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ggatcctttagtggtttacactgact 31  
 ||||| | | ||||| ||||| |||||  
 Db 36264 GGATCCTGTAGTGTTCACACTGACT 36291

## RESULT 6

AP000408/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

AP000408 269711 bp DNA linear HTG 28-JUL-2001  
 Homo sapiens chromosome 11 clone CMB9-40P16 map 11q13, WORKING  
 DRAFT SEQUENCE, 56 unordered pieces.  
 AP000408 GI:15027544  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 Homo sapiens DNA, clone:CMB9-40P16.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 269711)  
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 Published Only in Database (1999) In press  
 2 (bases 1 to 269711)  
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 Direct Submission  
 Submitted (27-AUG-1999), Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 On Jul 27, 2001 this sequence version replaced gi:8118747.  
 ----- Genome Center  
 Center: RIKEN Genomic Sciences Center (GSC)  
 Center code: RIKEN  
 Web site: http://hgp.gsc.riken.go.jp/  
 Contact: hattori@gsc.riken.go.jp  
 ----- Project Information  
 Center project name: Humbractfl  
 Center clone name: CMB9-40P16  
 ----- Summary Statistics  
 Sequencing vector: PCR products; 100% of reads  
 Chemistry: Dye-terminator ET-amersham; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 250673 bases at least Q40  
 Consensus quality: 258132 bases at least Q30  
 Consensus quality: 261704 bases at least Q20  
 Insert size: 264211; sum-of-contigs  
 Quality coverage: 8.26x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of  
 56 contigs. The true order of the pieces is not known and their  
 order in this sequence record is arbitrary. Gaps between the  
 contigs are represented as runs N, but the exact sizes of the gaps  
 are unknown. This record will be updated with the finished sequence  
 as soon as it is available and the accession number will be  
 preserved

1 19653 contig of 19653 bp in length  
 19754 39362 contig of 19609 bp in length  
 39463 55717 contig of 16255 bp in length  
 55818 67507 contig of 11690 bp in length  
 67608 77531 contig of 9924 bp in length  
 77632 87218 contig of 9587 bp in length  
 87319 95944 contig of 8626 bp in length  
 96045 100621 contig of 4577 bp in length

```

100722 113809 contig of 13088 bp in length
113910 117944 contig of 4035 bp in length
118045 121621 contig of 3577 bp in length
121722 126441 contig of 4720 bp in length
126542 130528 contig of 3987 bp in length
130629 138556 contig of 8228 bp in length
138957 146999 contig of 8043 bp in length
147100 150987 contig of 2845 bp in length
150988 156963 contig of 5029 bp in length
156964 165962 contig of 4477 bp in length
165963 172958 contig of 3935 bp in length
172959 180425 contig of 3217 bp in length
180426 211931 contig of 3806 bp in length
211932 215963 contig of 3912 bp in length
215964 217202 contig of 1139 bp in length
217203 220289 contig of 2987 bp in length
220290 222754 contig of 2365 bp in length
222755 225838 contig of 2365 bp in length
225839 228456 contig of 2984 bp in length
228457 231507 contig of 2951 bp in length
231508 234284 contig of 2677 bp in length
234285 236143 contig of 100 bp
236144 238375 contig of 2132 bp in length
238376 238475 contig of 100 bp
238476 240676 contig of 2201 bp in length
240677 242833 contig of 2057 bp in length
242834 242933 contig of 100 bp
242934 244625 contig of 1692 bp in length
244626 244725 contig of 100 bp
244726 245870 contig of 1145 bp in length
245871 245970 contig of 100 bp
245971 248069 contig of 2099 bp in length
248070 248169 contig of 100 bp
248170 250179 contig of 2010 bp in length
250180 250279 contig of 100 bp

```

NOTE: This is a 'working draft' sequence. It currently consists of 56 contigs, the true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 19653: contig of 19653 bp in length
19654 19753: gap of 100 bp
19754 39362: contig of 19609 bp in length
39363 39462: gap of 100 bp
39463 55717: contig of 16255 bp in length
55718 55817: gap of 100 bp
55818 67507: contig of 11690 bp in length
67508 67607: gap of 100 bp
67608 77531: contig of 9924 bp in length
77532 77631: gap of 100 bp
77632 87218: contig of 9587 bp in length
87219 87318: gap of 100 bp
87319 95944: contig of 8626 bp in length
95945 96044: gap of 100 bp
96045 100621: contig of 4577 bp in length
100622 100721: gap of 100 bp
100722 113809: contig of 13088 bp in length

```

```

113810 113909: gap of 100 bp
113910 117944: contig of 4035 bp in length
117945 118044: gap of 100 bp
118045 121621: contig of 3577 bp in length
121622 121721: gap of 100 bp
121722 126441: contig of 4720 bp in length
126442 126541: gap of 100 bp
126542 130528: contig of 3987 bp in length
130529 130628: gap of 100 bp
130629 138556: contig of 8228 bp in length
138557 138956: gap of 100 bp
138957 146999: contig of 8043 bp in length
147000 147099: gap of 100 bp
147100 150987: contig of 2845 bp in length
150988 156963: gap of 100 bp
156964 165962: contig of 4477 bp in length
165963 172958: contig of 3935 bp in length
172959 180425: contig of 3217 bp in length
180426 211931: contig of 3806 bp in length
211932 215963: contig of 3912 bp in length
215964 217202: contig of 1139 bp in length
217203 220289: contig of 2987 bp in length
220290 222754: contig of 2365 bp in length
222755 225838: contig of 2365 bp in length
225839 228456: contig of 2984 bp in length
228457 231507: contig of 2951 bp in length
231508 234284: contig of 2677 bp in length
234285 236143: contig of 100 bp
236144 238375: contig of 2132 bp in length
238376 238475: gap of 100 bp
238476 240676: contig of 2201 bp in length
240677 242833: contig of 2057 bp in length
242834 242933: gap of 100 bp
242934 244625: contig of 1692 bp in length
244626 244725: gap of 100 bp
244726 245870: contig of 1145 bp in length
245871 245970: gap of 100 bp
245971 248069: contig of 2099 bp in length
248070 248169: gap of 100 bp
248170 250179: contig of 2010 bp in length
250180 250279: gap of 100 bp

```

Query Match 70.3%; Score 23.2; DB 2; Length 269711;  
 Best Local Similarity 89.3%; Pred. No. 6.8;  
 Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 gatcctttatggtgtttacactgactgc 32  
|||||  
Db 172737 GATCCTTTATGATGTTTCTCTGACTGC 172710

RESULT 7  
A00633  
LOCUS A00633 33 bp DNA linear PAT 28-JAN-1993  
DEFINITION Nucleotide sequence 1 from patent number WO8603517.  
ACCESSION A00633  
VERSION A00633.1 GI:14563  
KEYWORDS unidentified.  
ORGANISM unidentified  
SOURCE unclassified.  
REFERENCE 1 (bases 1 to 33)  
AUTHORS  
TITLE DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR  
PREPARING SUCH PROTEIN  
JOURNAL Patent: WO 8603517-A 1 19-JUN-1986;  
FEATURES Location/Qualifiers  
source  
1..33  
/organism="unidentified"  
/db\_xref="taxon:32644" 11 t  
BASE COUNT 9 a 7 c 6 g 11 t  
ORIGIN

Query Match 67.9%; Score 22.4; DB 6; Length 33;  
Best Local Similarity 95.8%; Pred. No. 27;  
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 ttatggtgtttacactgactgc 33  
|||||  
Db 4 TCTATGTTGTTTACACTGACTGC 27

RESULT 8  
SYNTHIRMED  
LOCUS SYNTHIRMED 220 bp DNA linear SYN 27-APR-1993  
DEFINITION Synthetic hirudin gene, 3' end.  
ACCESSION M14964  
VERSION M14964.1 GI:208480  
KEYWORDS hirudin.  
SOURCE Synthetic DNA.  
ORGANISM synthetic construct  
artificial sequence.  
REFERENCE 1 (bases 1 to 220)  
AUTHORS Fortkamp, E., Rieger, M., Heisterberg-Moutses, G., Schweitzer, S. and Sommer, K.  
TITLE Cloning and expression in Escherichia coli of a synthetic DNA for  
hirudin, the blood coagulation inhibitor in the leech  
JOURNAL DNA 5, 511-517 (1986)  
MEDLINE 87132924  
COMMENT Draft entry and printed copy of sequence for [1] kindly provided by  
M. Rieger, 23-NOV-1988.  
FEATURES Location/Qualifiers  
source  
1..220  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
CDS  
1..207  
/note="hirudin precursor"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AA072556.1"  
/db\_xref="GI:208481"  
/translation="NSMVVYTDCTESGQNLCEGNSVCGQGNKILGSDGKKNOCVT  
GEGTPKQSHNDGDFEIPPEYLQ"  
mat\_peptide 10..204  
/product="hirudin"  
BASE COUNT 61 a 50 c 59 g 50 t  
ORIGIN EcoRI site.

Query Match 67.9%; Score 22.4; DB 12; Length 220;  
Best Local Similarity 95.8%; Pred. No. 24;  
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 ttatggtgtttacactgactgc 33  
|||||  
Db 4 TCTATGTTGTTTACACTGACTGC 27

RESULT 9  
A00630  
LOCUS A00630 224 bp DNA linear PAT 06-JUL-1995  
DEFINITION Artificial sequence for hirudin-like protein.  
ACCESSION A00630  
VERSION A00630.1 GI:14560  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 224)  
AUTHORS  
TITLE DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR  
PREPARING SUCH PROTEIN  
JOURNAL Patent: WO 8603517-A 15 19-JUN-1986;  
FEATURES Location/Qualifiers  
source  
1..224  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
CDS  
1..207  
/codon\_start=1  
/transl\_table=1  
/product="hirudin-like protein"  
/protein\_id="CAA00072.1"  
/db\_xref="GI:14561"  
/translation="NSMVVYTDCTESGQNLCEGNSVCGQGNKILGSDGKKNOCVT  
GEGTPKQSHNDGDFEIPPEYLQ"

BASE COUNT 62 a 51 c 60 g 51 t  
ORIGIN

Query Match 67.9%; Score 22.4; DB 6; Length 224;  
Best Local Similarity 95.8%; Pred. No. 24;  
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 ttatggtgtttacactgactgc 33  
|||||  
Db 4 TCTATGTTGTTTACACTGACTGC 27

RESULT 10  
A00631/c  
LOCUS A00631 224 bp DNA linear PAT 30-MAR-1993  
DEFINITION Artificial sequence for hirudin-like protein, reverse complement.  
ACCESSION A00631  
VERSION A00631.1 GI:14562  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 224)  
AUTHORS  
TITLE DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR  
PREPARING SUCH PROTEIN  
JOURNAL Patent: WO 8603517-A 16 19-JUN-1986;  
FEATURES Location/Qualifiers  
source  
1..224  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 51 a 60 c 51 g 62 t  
ORIGIN



AUTHORS . DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR  
 TITLE PREPARING SUCH PROTEIN  
 JOURNAL Patent: WO 8603517-A 2 19-JUN-1986;  
 FEATURES Location/Qualifiers  
 source 1..34  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 11 a 8 c 8 g 7 t  
 ORIGIN

Query Match 66.7%; Score 22; DB 6; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tatggtgtttacactgactgc 33  
 ||||||||||||||||||  
 Db 33 TATGTTGTTTACTGACTGC 12

Search completed: July 15, 2002, 22:29:14  
 Job time: 19000 sec

GenCore version 4.5  
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:49 ; Search time 758.37 Seconds  
(without alignments)  
74.710 Million cell updates/sec

Title: US-10-053-641-5

Perfect score: 33

Sequence: 1 tcggatcccttattggtttacactgactgc 33

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No.         | Score | Query Match | Length | DB ID | Description |
|--------------------|-------|-------------|--------|-------|-------------|
| 1                  | 22.4  | 67.9        | 224    | 7     | AA60746     |
| C 2                | 21.4  | 64.8        | 11472  | 23    | ABL24190    |
| 3                  | 21    | 63.6        | 726    | 12    | AAQ12379    |
| C 4                | 21    | 63.6        | 4809   | 22    | AA633924    |
| C 5                | 21    | 63.6        | 4809   | 22    | AAH93831    |
| 6                  | 20.8  | 63.0        | 217    | 7     | AA60355     |
| 7                  | 20.8  | 63.0        | 217    | 8     | AA60355     |
| 8                  | 20.2  | 61.2        | 9968   | 22    | AA636353    |
| 9                  | 19.6  | 59.4        | 36     | 19    | AA661848    |
| DNA encoding hirud |       |             |        |       |             |
| Drosophila melanog |       |             |        |       |             |
| Hirudin peptide/pr |       |             |        |       |             |
| Human prostate cDN |       |             |        |       |             |
| Human prostate-spe |       |             |        |       |             |
| Desulphatohirudin  |       |             |        |       |             |
| Sequence of the de |       |             |        |       |             |
| Human musculoskele |       |             |        |       |             |
| A. contortrix prot |       |             |        |       |             |

|      |      |      |        |    |          |                    |
|------|------|------|--------|----|----------|--------------------|
| C 10 | 19.4 | 58.8 | 572    | 22 | ABA14188 | Human nervous syst |
| C 11 | 19.4 | 58.8 | 2362   | 22 | AAH15864 | Human cDNA sequenc |
| C 12 | 19.4 | 58.8 | 3112   | 21 | AA658592 | Human PRO701 prote |
| C 13 | 19.4 | 58.8 | 3113   | 20 | AA234209 | Human PRO701 nucle |
| C 14 | 19.4 | 58.8 | 3113   | 21 | AA678552 | Human PRO701 (UNQ3 |
| C 15 | 19.4 | 58.8 | 3537   | 22 | ABA21445 | Human nervous syst |
| C 16 | 19.4 | 58.8 | 3539   | 22 | ABA21444 | Human nervous syst |
| C 17 | 19.4 | 58.8 | 10898  | 22 | AAK86165 | Human immune/haema |
| C 18 | 19.2 | 58.2 | 26767  | 23 | ABL08654 | Drosophila melanog |
| C 19 | 19.2 | 58.2 | 26960  | 23 | ABL08712 | Drosophila melanog |
| C 20 | 19   | 57.6 | 1323   | 22 | AAH73251 | Human cervical can |
| C 21 | 18.8 | 57.0 | 841    | 21 | AA652083 | Pigment protein fr |
| C 22 | 18.8 | 57.0 | 1193   | 21 | AA636135 | Arabidopsis thalia |
| C 23 | 18.8 | 57.0 | 1194   | 21 | AA646472 | Arabidopsis thalia |
| C 24 | 18.6 | 56.4 | 525    | 23 | AA658147 | cDNA #823 encoding |
| C 25 | 18.6 | 56.4 | 1846   | 22 | AAH15263 | Human cDNA sequenc |
| C 26 | 18.6 | 56.4 | 2837   | 23 | AA687012 | DNA encoding novel |
| C 27 | 18.6 | 56.4 | 3127   | 22 | AAH54412 | S. epidermidis gen |
| C 28 | 18.6 | 56.4 | 3157   | 22 | AAH54304 | S. epidermidis gen |
| C 29 | 18.6 | 56.4 | 17995  | 22 | AAK85152 | Human immune/haema |
| C 30 | 18.6 | 56.4 | 151826 | 21 | AAF22291 | BAC containing rep |
| C 31 | 18.4 | 55.8 | 300    | 21 | AAAO0372 | Human colon cancer |
| C 32 | 18.4 | 55.8 | 2006   | 21 | AA642229 | Arabidopsis thalia |
| C 33 | 18.4 | 55.8 | 4968   | 23 | ABL13204 | Drosophila melanog |
| C 34 | 18.2 | 55.2 | 557    | 20 | AA600474 | Human secreted pro |
| C 35 | 18.2 | 55.2 | 567    | 22 | AA630022 | Human lung antigen |
| C 36 | 18.2 | 55.2 | 567    | 22 | AA630023 | Human lung antigen |
| C 37 | 18.2 | 55.2 | 567    | 22 | AA628715 | Genomic sequence # |
| C 38 | 18.2 | 55.2 | 567    | 22 | AA628716 | Genomic sequence # |
| C 39 | 18.2 | 55.2 | 830    | 22 | AAK56695 | Human immune/haema |
| C 40 | 18.2 | 55.2 | 7263   | 23 | ABL23862 | Drosophila melanog |
| C 41 | 18.2 | 55.2 | 8143   | 22 | AA640435 | DNA encoding human |
| C 42 | 18.2 | 55.2 | 8143   | 22 | AA640435 | Human reproductive |
| C 43 | 18.2 | 55.2 | 32191  | 22 | AA626704 | Human genomic DNA  |
| C 44 | 18.2 | 55.2 | 49561  | 22 | AAK82012 | Human immune/haema |
| C 45 | 18   | 54.5 | 36     | 18 | AA645950 | Hirudin gene linke |

## ALIGNMENTS

RESULT 1  
AA60746  
ID AA60746 standard; DNA; 224 BP.  
XX  
AC AA60746;  
XX  
AC AA60746;  
XX  
DT 01-JAN-1980 (first entry)  
XX  
DE DNA encoding hirudin-like polypeptide.  
XX  
DE Hirudin; thrombin-antagonist; anticoagulant; ds.  
XX  
OS Synthetic.  
XX  
PN DE3445517-A.  
XX  
XX 19-JUN-1986.  
XX  
PF 13-DEC-1984; 84DE-3445517.  
XX  
PR 13-DEC-1984; 84DE-3445517.  
XX  
PA (GENB-) GEN-BIO-TEC GEN.  
XX  
PI Fortkamp E, Rieger M, Sommer R;  
XX  
DR WPI; 1986-162802/26.  
XX  
DR P-PSDB; AAP60827.  
XX  
PT New DNA sequence coding for new hirudin like polypeptide - useful  
XX as thrombin antagonist, e.g. for inhibiting blood coagulation

PS Disclosure; Fig. 1; 25pp; German.  
XX The sequence encodes a protein with hirudin-like activity. It is a  
CC thrombin-antagonist useful for inhibiting blood coagulation and for  
CC treating inflammation and/or oedema. It can be prepared in large  
CC amounts, unlike natural hirudin which is available only in tiny  
CC amounts from leeches. This polypeptide is produced using  
CC recombinant DNA techniques and is expressed in a bacterial  
CC transformant e.g. Escherichia coli.  
XX Sequence 224 BP; 61 A; 51 C; 60 G; 51 T; 1 other;  
SQ

Query Match 67.9%; Score 22.4; DB 7; Length 224;  
Best Local Similarity 95.8%; Pred. No. 3.6;  
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
0;

Qy 10 ttatggtgtttacactgactgc 33  
Db 4 tctatggtgtttacactgactgc 27

RESULT 2  
ABL24190/C  
ID ABL24190 standard; DNA; 11472 BP.  
XX  
AC ABL24190;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 24043.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 24043; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB27022).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 11472 BP; 3419 A; 2447 C; 2370 G; 3236 T; 0 other;

Query Match 64.8%; Score 21.4; DB 23; Length 11472;  
Best Local Similarity 80.6%; Pred. No. 16;  
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
0;

Qy 1 tcgggtatcttatggtgtttacactgactgc 31  
Db 10011 TCGGAGCGGTTATGTTGTTTACT 9981

RESULT 3  
AAQ12379  
ID AAQ12379 standard; DNA; 726 BP.  
XX  
AC AAQ12379;  
XX  
DT 17-SEP-1991 (first entry)  
XX  
DE Hirudin peptide/Protein A fusion gene.  
XX  
KW antioagulant; fusion protein; ds.  
XX  
FH Key Location/Qualifiers  
FT mat\_peptide 1..489  
FT /\*tag= a  
FT /product= Protein A  
FT 529..720  
FT mat\_peptide  
FT /\*tag= b  
FT /product= Hirudin  
FT 490..528  
FT misc\_RNA  
FT /\*tag= c  
FT /product= linker oligopeptide Y  
XX  
XX DE3942580-A.  
XX  
XX PN 27-JUN-1991.  
XX  
XX PD 22-DEC-1989; 89DE-3942580.  
XX  
XX PF 22-DEC-1989; 89DE-3942580.  
XX  
XX PR 22-DEC-1989; 89DE-3942580.  
XX  
XX PA (BADI ) BASF AG.  
XX  
XX PI Korwer W;  
XX  
XX DR WPI; 1991-194236/27.  
XX  
XX DR P-PSDB; AAR12751.  
XX  
XX PT Hirudin peptide prodn. by cleaving new fusion peptide - of  
XX PT hirudin and protein A, expressed in high yield and stable,  
XX PT soluble form by transformed E.coli  
XX  
XX PS Example 1; Page 6-7; 9pp; German.  
XX  
XX This sequence is an example of a fusion construct for expression of  
CC the fusion peptide of the invention. The Met residue in the linker  
CC oligopeptide allows cleavage by CNBr to release two fragments which  
CC can be easily separated by IgG affinity chromatography. (The protein  
CC a component binds to IgG sepharose). Increased yields of hirudin are  
CC obtained using this fusion construct.  
XX  
XX SQ Sequence 726 BP; 280 A; 171 C; 126 G; 149 T; 0 other;

Query Match 63.6%; Score 21; DB 12; Length 726;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
0;

Qy 13 atgtgtttacactgactgc 33  
Db 523 atgtgtttacactgactgc 543

RESULT 4





```
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT CDS 10..207
XX FT /*tag= a
XX FT /product= hirudin
XX PN EP168342-A.
XX PD 15-JAN-1986.
XX PF 10-JUN-1995; 85EP-0810268.
XX PR 14-JUN-1984; 84CH-0288284.
XX PA (CIBA ) CIBA GEIGY AG.
XX PI Liersch M, Rink H, Marki W, Grutter MG, Meyhack B;
XX DR WPI; 1986-015589/03.
XX DR P-PSDB; AAP60395.
XX PT DNA sequences coding for hirudin and derivs. - and expression
XX PT vectors, transformed cells, monoclonal antibodies and hybridomas,
XX PT useful as thrombin inhibitor.
XX PS Disclosure; Page 10; 123pp; German.
XX CC The sequence comprises an EcoRI restriction enzyme site at the
XX CC 5'-end and a BamHI restriction enzyme site at the 3'-end.
XX CC The sequence may be introduced into a vector for the transformation
XX CC of hosts, e.g. E. coli. The hirudin can thus be prepared on a
XX CC large scale. The product and its derivs. are thrombin inhibitors,
XX CC useful in anticoagulant therapy, esp. when injected at doses
XX CC of 0.01-0.05 mg/Kg.
XX SQ Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;

Query Match 63.0%; Score 20.8; DB 7; Length 217;
Best Local Similarity 91.7%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 ttatggtgtgtttacactgactgc 33
Db 7 ttcatggtgtttacacgactgc 30

RESULT 7
AAN70323
ID AAN70323 standard; DNA; 217 BP.
XX AC AAN70323;
XX DT 02-APR-1991 (first entry)
XX DE Sequence of the desulphatohirudin gene.
XX KW Anticoagulant; thrombin inhibitor; ds.
XX PN EP225633-A.
XX PD 16-JUN-1987.
XX PF 09-DEC-1986; 86EP-0117098.
XX PR 29-MAY-1986; 86GB-0013088.
XX PR 12-DEC-1985; 85GB-0030631.
XX PA (CIBA ) CIBA GEIGY AG.
XX PA (PLAN-) PLANTORGAN WERK HEINRICH.
XX PA (CHRI-) PLANTORGANW CHRISTENSEN.

XX PI Meyhack B, Marki W, Heim J;
XX DR WPI; 1987-164868/24.
XX PT New DNA constructs and hybrid vectors for transformation of yeast
XX PT etc. - useful for prodn. and secretion of protein with hirudin
XX PT activity for use as thrombin inhibitors.
XX PS Example; p44; 146pp; English.
XX CC The preferred DNA construct of the invention contains the PHO5
XX CC promoter and a DNA segment consisting of the PHO5 signal sequence
XX CC upstream of and in reading frame with a DNA sequence coding for
XX CC mature desulphatohirudin. The segment is under the transcriptional
XX CC control of the PHO5 promoter and the 3' flanking sequence of the
XX CC PHO5 gene.
XX SQ Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;

Query Match 63.0%; Score 20.8; DB 8; Length 217;
Best Local Similarity 91.7%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 ttatggtgtgtttacactgactgc 33
Db 7 ttcatggtgtttacacgactgc 30

RESULT 8
AAL36353
ID AAL36353 standard; DNA; 9968 BP.
XX AC AAL36353;
XX DT 08-JAN-2002 (first entry)
XX DE Human musculoskeletal system related polynucleotide SEQ ID NO 2718.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;
XX KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; human; secreted protein;
XX KW musculoskeletal system; ds.
XX OS Homo sapiens.
XX PN WO200155367-A1.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01338.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
```



CC and (f) infectious diseases such as viral, bacterial, fungal and  
 CC parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 9968 BP; 2324 A; 2302 C; 2695 G; 2647 T; 0 other;

Query Match 61.2%; Score 20.2; DB 22; Length 9968;  
 Best Local Similarity 88.0%; Pred. No. 52;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ggatccttattgtgtttactacg 28

Db 9864 ggatcgtttattgtttactacg 9888

# RESULT 9

AAV61848  
 ID AAV61848 standard; DNA; 36 BP.

XX AAV61848;

AC (first entry)

DT 11-DEC-1998

DE A. contortrix protein C activator PCR primer #3.

KW Protein C activator; anticoagulant; membrane-bound; inhibitor; clotting;  
 KW endothelial cell; organ rejection; transplantation; xenotransplantation.  
 KW PCR primer; ss.

XX Synthetic.

OS Agkistrodon contortrix.

XX WO9842850-A1.

PN 01-OCT-1998.

PD 26-MAR-1998; 98WO-GB00850.

PF 23-SEP-1997; 97GB-0020248.

PR 26-MAR-1997; 97GB-0006327.

XX (RPMS-) RPMS TECHNOLOGY LTD.

PI Dorling A, George AJT, Lechler RI, Riesbeck K;

XX WPI; 1998-557034/47.

PT New fusion proteins used particularly for preventing transplant  
 PT rejection - comprising a region with anticoagulant activity and a  
 PT region which can anchor the protein to a cell membrane

XX Disclosure; Page 29; 67pp; English.

CC AAV61846-V61860 are PCR primers used to amplify a protein C activator  
 CC from Agkistrodon contortrix. This protein contains a region with  
 CC anticoagulant activity and a region which can anchor the protein to a  
 CC cell membrane. The protein can provide membrane-bound anticoagulant  
 CC proteins. They are suitable for inhibiting the clotting cascade at the  
 CC surface of endothelial cells (ECs) thus inhibiting in vivo mechanisms  
 CC responsible for organ rejection. They can provide biological tissue  
 CC suitable for transplantation, particularly for xenotransplantation.

XX Sequence 36 BP; 7 A; 12 C; 9 G; 8 T; 0 other;

## Query Match

Best Local Similarity 59.4%; Score 19.6; DB 19; Length 36;  
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 cctttattgtttactacgactgc 33

Db 8 cctgggcgtgtttactacgactgc 33

## RESULT 10

ABAL4188/C

ID ABA14188 standard; cDNA; 572 BP.

XX ABA14188;

AC 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 3195.

KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
 KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.

XX Homo sapiens.

XX WO200159063-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180828.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226868.

XX 22-AUG-2000; 2000US-0227182.

XX 23-AUG-2000; 2000US-0227009.

XX 30-AUG-2000; 2000US-0228924.

XX 01-SEP-2000; 2000US-0229287.

XX 01-SEP-2000; 2000US-0229343.

XX 01-SEP-2000; 2000US-0229344.

XX 01-SEP-2000; 2000US-0229345.

XX 05-SEP-2000; 2000US-0229509.

XX 05-SEP-2000; 2000US-0229513.

XX 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0231967.  
PR 14-SEP-2000; 2000US-0232396.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241178.  
PR 20-OCT-2000; 2000US-0241186.  
PR 20-OCT-2000; 2000US-0241187.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2000US-0254097.  
PR 05-JAN-2001; 2000US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPT; 2001-541565/60.  
DR P-PSDB; ABB17862.  
XX

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -

Claim 1; SEQ ID NO 3195; 1701pp + Sequence Listing; English.

The invention relates to novel genes (ABAI1004-ABA21534) and proteins (ABAI14678-ABAI18001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 572 BP; 171 A; 169 C; 106 G; 119 T; 7 other;

Query Match 58.8%; Score 19.4; DB 22; Length 572;  
Best Local Similarity 79.3%; Pred No. 78;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 ggatccttattggtgtttacactgactg 32  
||| |||| |||| |||| |||| ||||  
Db 223 GGAACCTTTGTTGTTGTAACACTACTG 195

RESULT 11  
AAH15864/C  
ID AAH15864 standard; CDNA; 2362 BP.  
XX  
AC AAH15864;  
XX  
DT 26-JUN-2001 (first entry)  
XX

Human cDNA sequence SEQ ID NO:14380.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEQ ID 14380; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialized methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2362 BP; 654 A; 671 C; 534 G; 503 T; 0 other;

Query Match 58.8%; Score 19.4; DB 22; Length 2362;

Best Local Similarity 79.3%; Pred. NO. 94;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 ggatccttattggtgttttaactgactg 32

||||| ||||| ||||| ||||| |||||

Db 1139 GGAACCTTTGTGTTGTTGAACATACATG 1111

RESULT 12

AAAC58592/C

ID AAC58592 standard; cDNA; 3112 BP.

XX AAC58592;

XX

29-JAN-2001 (first entry)

Human PRO701 protein UNQ365 encoding cDNA SEQ ID NO:66.

Human; immune related disease; diagnosis; antiinflammatory; cardiant; dermatological; antiarthritic; antirheumatic; immunosuppressive; haemostatic; antithyroid; antidiabetic; nootropic; neuroprotective; antianaemic; hepatotropic; virucide; antipsoriatic; antiallergic; antiasthmatic; systemic lupus erythematosus; rheumatoid arthritis; osteoarthritis; spondyloarthritis; systemic sclerosis; sarcoidosis; idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis; systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus; autoimmune thrombocytopaenia; immune-mediated renal disease; demyelinating disease; hepatobiliary disease; Whipple's disease; inflammatory bowel disease; gluten-sensitive enteropathy; autoimmune disease; immune-mediated skin disease; allergic disease; immunological disease; transplantation associated disease; graft rejection; graft-versus-host-disease; ss.

Homo sapiens.

WO200053758-A2.

14-SEP-2000.

02-MAR-2000; 2000WO-US05841.

08-MAR-1999; 99WO-US05028.

10-MAR-1999; 99US-0123618.

12-MAR-1999; 99US-0123957.

23-MAR-1999; 99US-0125775.

12-APR-1999; 99US-0128849.

20-APR-1999; 99WO-US08615.

28-APR-1999; 99US-0131445.

04-MAY-1999; 99US-0132371.

14-MAY-1999; 99US-0134287.

02-JUN-1999; 99WO-US12252.

23-JUN-1999; 99US-0141037.

20-JUL-1999; 99US-0144758.

26-JUL-1999; 99US-0145698.

28-JUL-1999; 99US-0146222.

01-SEP-1999; 99WO-US20111.

08-SEP-1999; 99WO-US20594.

13-SEP-1999; 99WO-US20944.

15-SEP-1999; 99WO-US21090.

15-SEP-1999; 99WO-US21547.

05-OCT-1999; 99WO-US23089.

29-OCT-1999; 99US-0162506.

30-NOV-1999; 99WO-US28214.

30-NOV-1999; 99WO-US28313.

30-NOV-1999; 99WO-US28409.

01-DEC-1999; 99WO-US28301.

01-DEC-1999; 99WO-US28634.

02-DEC-1999; 99WO-US28551.

02-DEC-1999; 99WO-US28564.

02-DEC-1999; 99WO-US28565.

16-DEC-1999; 99WO-US30095.

20-DEC-1999; 99WO-US30999.

30-DEC-1999; 99WO-US31274.

05-JAN-2000; 2000WO-US00219.

06-JAN-2000; 2000WO-US00277.

06-JAN-2000; 2000WO-US00376.

11-FEB-2000; 2000WO-US03565.

18-FEB-2000; 2000WO-US04341.

18-FEB-2000; 2000WO-US04342.

22-FEB-2000; 2000WO-US04414.

(GETH ) GENENTECH INC.

Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W;

Kabakoff RC, Lu Y, Pan J, Pennica D, Shelton DL, Smith V;

Stewart TA, Tumas D, Watanabe CK, Wood WI, Yan M;

XX

DR WPI: 2000-572271/53.  
 DR P-PSDB; AAB33427.  
 XX  
 PT Sixty four PRO polypeptides, useful in the diagnosis and treatment of  
 PT immune related disorders, e.g. systemic lupus erythematosus, rheumatoid  
 PT arthritis, osteoarthritis, thyroiditis and diabetes mellitus -  
 XX  
 PS Claim 23; Fig 27; 309pp; English.  
 XX  
 CC The present invention describes sixty four human PRO proteins which can  
 CC be used in the treatment of immune related diseases. The human PRO  
 CC proteins, anti-PRO antibodies, agonists and antagonists are useful for  
 CC treating and diagnosing immune related disorders. The disorders are  
 CC selected from systemic lupus erythematosus, rheumatoid arthritis,  
 CC osteoarthritis, juvenile chronic arthritis, spondyloarthropathies,  
 CC systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's  
 CC syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic  
 CC anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes mellitus,  
 CC immune-mediated renal disease, demyelinating diseases of the central  
 CC and peripheral nervous systems, hepatobiliary diseases, inflammatory  
 CC bowel disease, gluten-sensitive enteropathy and Whipple's disease,  
 CC autoimmune or immune-mediated skin diseases, allergic diseases,  
 CC immunological diseases of the lung, and transplantation associated  
 CC diseases including graft rejection and graft-versus-host-disease.  
 CC AAC58397 to AAC58578 represent PCR primers and hybridisation probes used  
 CC in the isolation of human PRO sequences. AAC58579 to AAC58642 and  
 CC AAB33414 to AAB33477 represent human PRO polynucleotide and protein  
 CC sequences given in the exemplification of the present invention.  
 XX  
 SQ Sequence 3112 BP; 853 A; 867 C; 712 G; 580 T; 0 other;

Query Match 58.8%; Score 19.4; DB 21; Length 3112;  
 Best Local Similarity 79.3%; Pred. No. 97;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 4 ggatcctttatgtgtgttttaactgactg 32  
 ||| ||||| ||||| ||| ||||  
 DB 1890 GGAACCTTTGTTGGTTGTAACATACG 1862

## RESULT 13

AAZ34209/C  
 ID AAZ34209 standard; cDNA; 3113 BP.

XX  
 AC AAZ34209;

XX  
 DT 07-DEC-1999 (first entry)

XX  
 DE Human PRO701 nucleotide sequence.

XX  
 KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
 KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
 KW secreted protein; transmembrane protein; ss.

XX OS Homo sapiens.

XX PN W09946281-A2.

XX PD 16-SEP-1999.

XX PF 08-MAR-1999; 99WO-US05028.

XX PR 10-MAR-1998; 98US-0077450.

PR 11-MAR-1998; 98US-0077632.

PR 11-MAR-1998; 98US-0077641.

PR 11-MAR-1998; 98US-0077649.

PR 12-MAR-1998; 98US-0077791.

PR 13-MAR-1998; 98US-0078004.

PR 17-MAR-1998; 98US-0040220.

PR 20-MAR-1998; 98US-0078886.

PR 20-MAR-1998; 98US-0078910.

PR 20-MAR-1998; 98US-0078936.

PR 20-MAR-1998; 98US-0078939.  
 PR 25-MAR-1998; 98US-0079294.  
 PR 26-MAR-1998; 98US-0079656.  
 PR 27-MAR-1998; 98US-0079663.  
 PR 27-MAR-1998; 98US-0079664.  
 PR 27-MAR-1998; 98US-0079689.  
 PR 27-MAR-1998; 98US-0079728.  
 PR 27-MAR-1998; 98US-0079786.  
 PR 30-MAR-1998; 98US-0079920.  
 PR 30-MAR-1998; 98US-0079923.  
 PR 31-MAR-1998; 98US-0080105.  
 PR 31-MAR-1998; 98US-0080107.  
 PR 31-MAR-1998; 98US-0080165.  
 PR 31-MAR-1998; 98US-0080194.  
 PR 01-APR-1998; 98US-0080327.  
 PR 01-APR-1998; 98US-0080328.  
 PR 01-APR-1998; 98US-0080333.  
 PR 01-APR-1998; 98US-0080334.  
 PR 08-APR-1998; 98US-0081049.  
 PR 08-APR-1998; 98US-0081070.  
 PR 08-APR-1998; 98US-0081071.  
 PR 09-APR-1998; 98US-0081195.  
 PR 09-APR-1998; 98US-0081203.  
 PR 09-APR-1998; 98US-0081229.  
 PR 15-APR-1998; 98US-0081817.  
 PR 15-APR-1998; 98US-0081838.  
 PR 15-APR-1998; 98US-0081952.  
 PR 15-APR-1998; 98US-0081955.  
 PR 21-APR-1998; 98US-0082568.  
 PR 21-APR-1998; 98US-0082569.  
 PR 22-APR-1998; 98US-0082700.  
 PR 22-APR-1998; 98US-0082704.  
 PR 23-APR-1998; 98US-0082804.  
 PR 23-APR-1998; 98US-0082767.  
 PR 23-APR-1998; 98US-0082796.  
 PR 27-APR-1998; 98US-0083336.  
 PR 28-APR-1998; 98US-0083322.  
 PR 29-APR-1998; 98US-0083392.  
 PR 29-APR-1998; 98US-0083495.  
 PR 29-APR-1998; 98US-0083496.  
 PR 29-APR-1998; 98US-0083499.  
 PR 29-APR-1998; 98US-0083500.  
 PR 29-APR-1998; 98US-0083545.  
 PR 29-APR-1998; 98US-0083554.  
 PR 29-APR-1998; 98US-0083558.  
 PR 29-APR-1998; 98US-0083559.  
 PR 30-APR-1998; 98US-0083742.  
 PR 05-MAY-1998; 98US-0084366.  
 PR 06-MAY-1998; 98US-0084414.  
 PR 06-MAY-1998; 98US-0084441.  
 PR 07-MAY-1998; 98US-0084598.  
 PR 07-MAY-1998; 98US-0084600.  
 PR 07-MAY-1998; 98US-0084627.  
 PR 07-MAY-1998; 98US-0084637.  
 PR 07-MAY-1998; 98US-0084639.  
 PR 07-MAY-1998; 98US-0084640.  
 PR 07-MAY-1998; 98US-0084643.  
 PR 13-MAY-1998; 98US-0085323.  
 PR 13-MAY-1998; 98US-0085338.  
 PR 13-MAY-1998; 98US-0085339.  
 PR 15-MAY-1998; 98US-0085573.  
 PR 15-MAY-1998; 98US-0085579.  
 PR 15-MAY-1998; 98US-0085580.  
 PR 15-MAY-1998; 98US-0085582.  
 PR 15-MAY-1998; 98US-0085689.  
 PR 15-MAY-1998; 98US-0085697.  
 PR 15-MAY-1998; 98US-0085700.  
 PR 15-MAY-1998; 98US-0085704.  
 PR 18-MAY-1998; 98US-0086023.  
 PR 22-MAY-1998; 98US-0086392.  
 PR 22-MAY-1998; 98US-0086414.  
 PR 22-MAY-1998; 98US-0086430.  
 PR 22-MAY-1998; 98US-0086486.

PR 28-MAY-1998; 98US-0087098.  
 PR 28-MAY-1998; 98US-0087106.  
 PR 28-MAY-1998; 98US-0087208.  
 PR 30-JUL-1998; 98US-0094651.  
 PR 11-SEP-1998; 98US-0100038.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX Wood WL, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;  
 XX WPI: 1999-551358/46.  
 DR P-PSDB; AAY41740.  
 XX  
 PT New secreted and transmembrane polypeptides and their polynucleotides,  
 PT useful for treating blood coagulation disorders, cancers and cellular  
 PT adhesion disorders -  
 XX  
 XX Claim 2; Fig 150; 530pp; English.  
 PS  
 CC The present invention describes secreted and transmembrane polypeptides  
 CC and their polynucleotides. The nucleotide sequences are useful as  
 CC sources of probes, primers, for chromosome mapping, and for generation  
 CC of antisense sequences. They can also be used to create transgenic  
 CC animals. The proteins can be used to treat a variety of diseases and  
 CC disorders, depending on their function. Diseases that may be treated  
 CC include blood coagulation disorders, cancers and cellular adhesion  
 CC disorders. They may also be used to raise antibodies. AAZ33891 to  
 CC AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and  
 CC polypeptide sequence given in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 3113 BP; 853 A; 867 C; 712 G; 681 T; 0 other;  
 Query Match 58.8%; Score 19.4; DB 20; Length 3113;  
 Best Local Similarity 79.3%; Pred. No. 97;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 4 ggatccttattggtgtttacactgactg 32  
 ||| ||||| ||||| || |||||  
 Db 1890 GGAACCTTTGTTGTTGTTGAACATACCTG 1862

RESULT 14  
 AAC78552/G  
 ID AAC78552 standard; cDNA; 3113 BP.  
 XX  
 AC AAC78552;  
 XX  
 DT 08-FEB-2001 (first entry)  
 DE Human PRO701 (UNQ365) nucleotide sequence SEQ ID NO:374.  
 XX  
 XX Human; secreted protein; transmembrane protein; PRO; EST; cytotstatic;  
 KW expressed sequence tag; detection; cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200053756-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PF 18-FEB-2000; 2000WO-US04341.  
 PR  
 PR 08-MAR-1999; 99WO-US05028.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 29-MAR-1999; 99US-0126773.  
 PR 21-APR-1999; 99US-0130232.  
 PR 28-APR-1999; 99US-0131445.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 30-DEC-1999; 99WO-US31243.  
 PR 30-DEC-1999; 99WO-US31274.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 XX  
 XX (GETH ) GENENTECH INC.  
 PA  
 XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;  
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
 PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;  
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WL;  
 XX  
 DR WPI: 2000-611443/58.  
 DR P-PSDB; AAB44296.  
 DR  
 XX Novel PRO polypeptides and polynucleotides used in detection methods,  
 PT to target bioactive molecules to specific cells, and to modulate  
 PT cellular activities -  
 XX  
 PS Claim 2; Fig 150; 636pp; English.  
 CC  
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed  
 CC sequence tag) sequences which encode secreted or transmembrane PRO  
 CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic  
 CC activity. The polynucleotides and polypeptides can be used for detecting  
 CC the presence of PRO polypeptides in samples, for linking bioactive  
 CC molecules to cells and for modulating biological activities of cells,  
 CC using the polypeptides for specific targeting. The polypeptide targeting  
 CC can be used to kill the target cells, e.g. for the treatment of cancers.  
 CC The polypeptide pairs provide specific targeting of bioactive molecules  
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in  
 CC the isolation of the PRO polynucleotide sequences.  
 XX  
 SQ Sequence 3113 BP; 853 A; 867 C; 712 G; 681 T; 0 other;  
 Query Match 58.8%; Score 19.4; DB 21; Length 3113;  
 Best Local Similarity 79.3%; Pred. No. 97;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 4 ggatccttattggtgtttacactgactg 32  
 ||| ||||| ||||| || |||||  
 Db 1890 GGAACCTTTGTTGTTGTTGAACATACCTG 1862

RESULT 15  
 ABA21445/c  
 ID ABA21445 standard; DNA; 3537 BP.  
 XX  
 AC ABA21445;  
 XX  
 DT 23-JAN-2002 (first entry)  
 DE  
 DE Human nervous system related polynucleotide SEQ ID NO 13776.  
 XX  
 XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
 KW antiparkinsonian; antisickling; antianemic; antiarthritic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
 OS Homo sapiens.  
 XX  
 PN WO200159063-A2.  
 XX



PD 16-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01334.  
PF 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226273.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226869.  
PR 22-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PR (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
XX  
DR

XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX  
XX Disclosure: SEQ ID NO 13776; 1701pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins  
CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 3537 BP; 1142 A; 722 C; 649 G; 1024 T; 0 other;

Query Match 58.8%; Score 19.4; DB 22; Length 3537;  
Best Local Similarity 79.3%; Pred. No. 99;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 4 ggatcctttatggtgtttacactgactg 32  
||| ||||| ||||| || |||||  
Db 223 GGAACCTTGTGTGTGTGAACATACTG 195

Search completed: July 15, 2002, 22:44:53  
Job time: 10066 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:31:49 ; Search time 174.99 Seconds  
(without alignments)  
46.322 Million cell updates/sec

Title: US-10-053-641-5  
Perfect score: 33  
Sequence: 1 tcgggaccccttattggtttacactgactgc 33

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 303533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 22    | 66.7        | 93     | 1     | US-07-859-453E-3  |
| 2          | 21    | 63.6        | 106    | 1     | US-08-262-384A-6  |
| 3          | 21    | 63.6        | 129    | 1     | US-08-262-384A-3  |
| 4          | 21    | 63.6        | 235    | 1     | US-08-262-384A-7  |
| 5          | 21    | 63.6        | 235    | 1     | US-08-262-384A-8  |
| 6          | 21    | 63.6        | 726    | 1     | US-08-262-384A-9  |
| 7          | 21    | 63.6        | 726    | 1     | US-08-262-384A-10 |
| 8          | 20.8  | 63.0        | 58     | 6     | 5422249-10        |
| 9          | 20.8  | 63.0        | 109    | 6     | 5422249-6         |
| 10         | 20.8  | 63.0        | 217    | 6     | 5422249-13        |
| 11         | 20.8  | 63.0        | 218    | 6     | 5422249-4         |
| 12         | 19.4  | 58.8        | 122    | 1     | US-08-619-598-4   |
| 13         | 18.6  | 56.4        | 5676   | 2     | US-08-663-998-3   |
| 14         | 18.6  | 56.4        | 5682   | 2     | US-08-663-998-4   |
| 15         | 18.4  | 55.8        | 41     | 1     | US-07-854-596B-50 |
| 16         | 18.4  | 55.8        | 43     | 1     | US-08-044-506B-9  |
| 17         | 18.4  | 55.8        | 44     | 1     | US-07-854-596B-51 |
| 18         | 18.4  | 55.8        | 36     | 1     | US-08-466-265-12  |
| 19         | 18.4  | 55.8        | 38     | 1     | US-08-466-265-11  |
| 20         | 18.4  | 55.8        | 36     | 1     | US-08-605-053-4   |
| 21         | 18.4  | 55.8        | 105    | 1     | US-07-842-089E-24 |
| 22         | 18.4  | 55.8        | 105    | 1     | US-08-264-485-24  |
| 23         | 17.8  | 53.9        | 227    | 6     | 5164304-3         |
| 24         | 17.8  | 53.9        | 228    | 6     | 5179196-3         |
| 25         | 17.4  | 52.7        | 39     | 1     | US-08-044-506B-8  |
| 26         | 17.4  | 52.7        | 1107   | 1     | US-08-292-945-5   |
| 27         | 17.4  | 52.7        | 1107   | 1     | US-08-252-073A-5  |

|   |    |      |      |      |   |                   |                    |
|---|----|------|------|------|---|-------------------|--------------------|
| c | 28 | 17.4 | 52.7 | 1107 | 5 | PCT-US93-12074-5  | Sequence 5, Appl   |
| c | 29 | 17.2 | 52.1 | 36   | 4 | US-08-235-836C-28 | Sequence 28, Appl  |
| c | 30 | 17.2 | 52.1 | 1794 | 3 | US-08-920-270-1   | Sequence 1, Appl   |
| c | 31 | 17.2 | 52.1 | 2203 | 1 | US-08-076-091C-13 | Sequence 13, Appl  |
| c | 32 | 17.2 | 52.1 | 2203 | 1 | US-08-283-641-13  | Sequence 13, Appl  |
| c | 33 | 17.2 | 52.1 | 5055 | 4 | US-09-242-632A-13 | Sequence 13, Appl  |
| c | 34 | 17.2 | 52.1 | 2772 | 3 | US-08-936-135-1   | Sequence 1, Appl   |
| c | 35 | 16.8 | 50.9 | 201  | 6 | 5422249-3         | Patent No. 5422249 |
| c | 36 | 16.6 | 50.3 | 736  | 4 | US-09-268-364-19  | Sequence 19, Appl  |
| c | 37 | 16.6 | 50.3 | 1131 | 1 | US-07-959-946-2   | Sequence 2, Appl   |
| c | 38 | 16.6 | 50.3 | 1131 | 1 | US-08-333-577-2   | Sequence 2, Appl   |
| c | 39 | 16.6 | 50.3 | 1131 | 5 | PCT-US92-08634-2  | Sequence 2, Appl   |
| c | 40 | 16.6 | 50.3 | 2473 | 2 | US-08-540-804-3   | Sequence 3, Appl   |
| c | 41 | 16.6 | 50.3 | 2473 | 2 | US-08-218-265-3   | Sequence 3, Appl   |
| c | 42 | 16.6 | 50.3 | 2473 | 3 | US-08-521-872-3   | Sequence 3, Appl   |
| c | 43 | 16.6 | 50.3 | 2473 | 4 | US-08-590-399-3   | Sequence 3, Appl   |
| c | 44 | 16.6 | 50.3 | 3147 | 2 | US-07-952-853-23  | Sequence 23, Appl  |
| c | 45 | 16.6 | 50.3 | 3147 | 2 | US-08-914-848-23  | Sequence 23, Appl  |

ALIGNMENTS

RESULT 1  
US-07-859-453E-3  
; Sequence 3, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:  
; APPLICANT: Kurfuerst, Manfred  
; APPLICANT: Ruebsamen, Klaus  
; APPLICANT: Schmied, Bernhard  
; APPLICANT: Koerwer, Wolfgang  
; APPLICANT: Schwenen, Juergen  
; APPLICANT: Hoeftken, Hans Wolfgang  
; TITLE OF INVENTION: Hirudin/polyalkylene glycol  
; TITLE OF INVENTION: conjugates  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kell & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS version 7.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/859,453E  
; FILING DATE: 29-MAY-1992  
; CLASSIFICATION: 514  
; CLASSIFICATION: A 61 K 37/64  
; CLASSIFICATION: A 61 K 47/48  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/01998  
; FILING DATE: 22-NOV-1990  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
US-07-859-453E-3

Query Match 66.7%; Score 22; DB 1; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.25;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tatgggtgtttacactgactgc 33  
|||||

Tue Jul 16 08:24:11 2002

```
Db 15 TATGTTGTTTACTGACTGC 36

RESULT 2
US-08-262-384A-6/c
; Sequence 6, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-262-384A-3

Query Match 63.6%; Score 21; DB 1; Length 129;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 atggtgttttacactgactgc 33
Db 31 ATGTTGTTTACTGACTGC 51

RESULT 4
US-08-262-384A-7
; Sequence 7, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-262-384A-7

Query Match 63.6%; Score 21; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 atggtgttttacactgactgc 33
Db 81 ATGTTGTTTACTGACTGC 61

RESULT 3
US-08-262-384A-3
; Sequence 3, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
```

Query Match 63.6%; Score 21; DB 1; Length 235;  
Best Local Similarity 100.0%; Pred. No. 0.96;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 atgggtgtttacactgactgc 33  
|||||  
Db 31 ATGGTGTGTTACTGACTGTC 51

RESULT 5  
US-08-262-384A-8/c  
Sequence 8, Application US/08262384A  
Patent No. 5624822  
GENERAL INFORMATION:  
APPLICANT: Koerwer, Wolfgang  
TITLE OF INVENTION: The Preparation of Hirudin  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Keil & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM AT-compatible, 80486 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/262,384A  
FILING DATE: 20-JUN-1994  
CLASSIFICATION: C 12 N 15/62  
CLASSIFICATION: C 12 N 15/31  
CLASSIFICATION: C 07 K 7/10  
CLASSIFICATION: C 12 P 21/02  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP90/02084  
FILING DATE: 04-DEC-1990  
APPLICATION NUMBER: US 07861820  
FILING DATE: 18-JUN-1992

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-262-384A-8

Query Match 63.6%; Score 21; DB 1; Length 235;  
Best Local Similarity 100.0%; Pred. No. 0.96;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 atgggtgtttacactgactgc 33  
|||||  
Db 209 ATGGTGTGTTACTGACTGTC 189

RESULT 6  
US-08-262-384A-9  
Sequence 9, Application US/08262384A  
Patent No. 5624822  
GENERAL INFORMATION:  
APPLICANT: Koerwer, Wolfgang  
TITLE OF INVENTION: The Preparation of Hirudin  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Keil & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington

STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM AT-compatible, 80486 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/262,384A  
FILING DATE: 20-JUN-1994  
CLASSIFICATION: C 12 N 15/62  
CLASSIFICATION: C 12 N 15/31  
CLASSIFICATION: C 07 K 7/10  
CLASSIFICATION: C 12 P 21/02  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP90/02084  
FILING DATE: 04-DEC-1990  
APPLICATION NUMBER: US 07861820  
FILING DATE: 18-JUN-1992

INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-262-384A-9

Query Match 63.6%; Score 21; DB 1; Length 726;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 atgggtgtttacactgactgc 33  
|||||  
Db 523 ATGGTGTGTTACTGACTGTC 543

RESULT 7  
US-08-262-384A-10/c  
Sequence 10, Application US/08262384A  
Patent No. 5624822  
GENERAL INFORMATION:  
APPLICANT: Koerwer, Wolfgang  
TITLE OF INVENTION: The Preparation of Hirudin  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Keil & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM AT-compatible, 80486 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/262,384A  
FILING DATE: 20-JUN-1994  
CLASSIFICATION: C 12 N 15/62  
CLASSIFICATION: C 12 N 15/31  
CLASSIFICATION: C 07 K 7/10  
CLASSIFICATION: C 12 P 21/02  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP90/02084  
FILING DATE: 04-DEC-1990  
APPLICATION NUMBER: US 07861820  
FILING DATE: 18-JUN-1992

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-262-384A-10



Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 10 ttatggtgtttacactgactgc 33  
|| ||||| ||||| ||||| |||||  
Db 7 ttcattgtgtttacaccgactgc 30

## RESULT 12

US-08-619-598-4/c  
; Sequence 4, Application US/08619598  
; Patent No. 5672487

## ; GENERAL INFORMATION:

; APPLICANT: SCHWEDEN, Juergen  
; APPLICANT: BOLLSCHEWILLER, Claus  
; APPLICANT: PIONTEK, Michael  
; APPLICANT: WEYDEMANN, Ulrike

; APPLICANT: GELLISSEN, Gerd

; TITLE OF INVENTION: THE RECOMBINANT PRODUCTION

; TITLE OF INVENTION: OF PROTEINS IN YEAST

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Keil & Weinlauf

; STREET: 1101 Connecticut Avenue

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20036

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage

; COMPUTER: IBM AT-compatible, 80486 processor

; OPERATING SYSTEM: MS-DOS version 6.0

; SOFTWARE: WordPerfect version 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/619,598

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP/94/03409

; FILING DATE: 17-OCT-1994

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 122 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

US-08-619-598-4

Query Match 58.8%; Score 19.4; DB 1; Length 122;

Best Local Similarity 95.2%; Pred. No. 4.4;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 13 atggtgtgtttacactgactgc 33

||||| ||||| ||||| |||||

Db 35 AGGCTTCTTACACTGACTGC 15

## RESULT 13

US-08-663-998-3/c

; Sequence 3, Application US/08663998

; Patent No. 5846946

## ; GENERAL INFORMATION:

; APPLICANT: HUEBNER, Robert C.

; APPLICANT: NORMAN, Jon A.

; APPLICANT: LIANG, Xiaowu

; APPLICANT: CARNER, Kristin R.

; APPLICANT: BARBOUR, Alan G.

; APPLICANT: LUKE, Catherine J.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

; TITLE OF INVENTION: ADMINISTERING BORRELIA DNA

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtis, Morris & Safford, P.C.  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/663,998  
; FILING DATE: 06-JUN-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kowalski, Thomas J.  
; REGISTRATION NUMBER: 32,147  
; REFERENCE/DOCKET NUMBER: 454312-2440  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-840-3333  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5676 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-663-998-3

Query Match 56.4%; Score 18.6; DB 2; Length 5676;

Best Local Similarity 72.7%; Pred. No. 22;

Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 tcgggatccttatggtgtttacactgactgc 33

||||| ||||| ||||| |||||

Db 2448 TCTGGATCCTTAAGGTTTTTTTGGACTTCTGC 2416

## RESULT 14

US-08-663-998-4/c

; Sequence 4, Application US/08663998

; Patent No. 5846946

## ; GENERAL INFORMATION:

; APPLICANT: HUEBNER, Robert C.

; APPLICANT: NORMAN, Jon A.

; APPLICANT: LIANG, Xiaowu

; APPLICANT: CARNER, Kristin R.

; APPLICANT: BARBOUR, Alan G.

; APPLICANT: LUKE, Catherine J.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

; TITLE OF INVENTION: ADMINISTERING BORRELIA DNA

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtis, Morris & Safford, P.C.

; STREET: 530 Fifth Avenue

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10036

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/663,998

; FILING DATE: 06-JUN-1996

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Kowalski, Thomas J.

; REGISTRATION NUMBER: 32,147

```

; REFERENCE/DOCKET NUMBER: 454312-2440
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: 212-840-3333
;
; INFORMATION FOR SEQ ID NO: 4:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 5682 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: DNA (genomic)
;
; US-08-663-998-4

```

Query Match 56.4%; Score 18.6; DB 2; Length 5682;  
Best Local Similarity 72.7%;  
Pred. No. 22;  
Matches 24; Conservative 0; Mismatches 9; Indels 0

**Qy**      1    tcgggacccctttatggttgattacactgactgc    33  
         || | | | | | | | | | | | | | | | |  
**Dd**     2454   TCTGGATCCTTAAGGTTTTTTTGGACTTTCTGC    2422

```

RESULT 15
US-07-854-596B-50
: Sequence 50, Application US/07854596B
: Patent No. 5434073
: GENERAL INFORMATION:
: APPLICANT: Dawson, Keith M
: APPLICANT: Hunter, Michael G
: APPLICANT: Czaplowski, Lloyd G
: TITLE OF INVENTION: Proteins and nucleic acids
: NUMBER OF SEQUENCES: 73
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dr. John J. McDonnell
: STREET: Ten South Wacker Drive, Suite 3000
: CITY: Chicago
: STATE: IL
: COUNTRY: USA
: ZIP: 60606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/854,596B
: FILING DATE: 03-JUN-1992
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: McDonnell, John J
: REGISTRATION NUMBER: 26,949
: REFERENCE/DOCKET NUMBER: 92,337
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-715-1000
: TELEFAX: 312-715-1234
: TELEX: 910-221-5317
: INFORMATION FOR SEQ ID NO: 50:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 41 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1..41
: OTHER INFORMATION: /note="oligonucleotide BB2"
US-07-854-596B-50

```

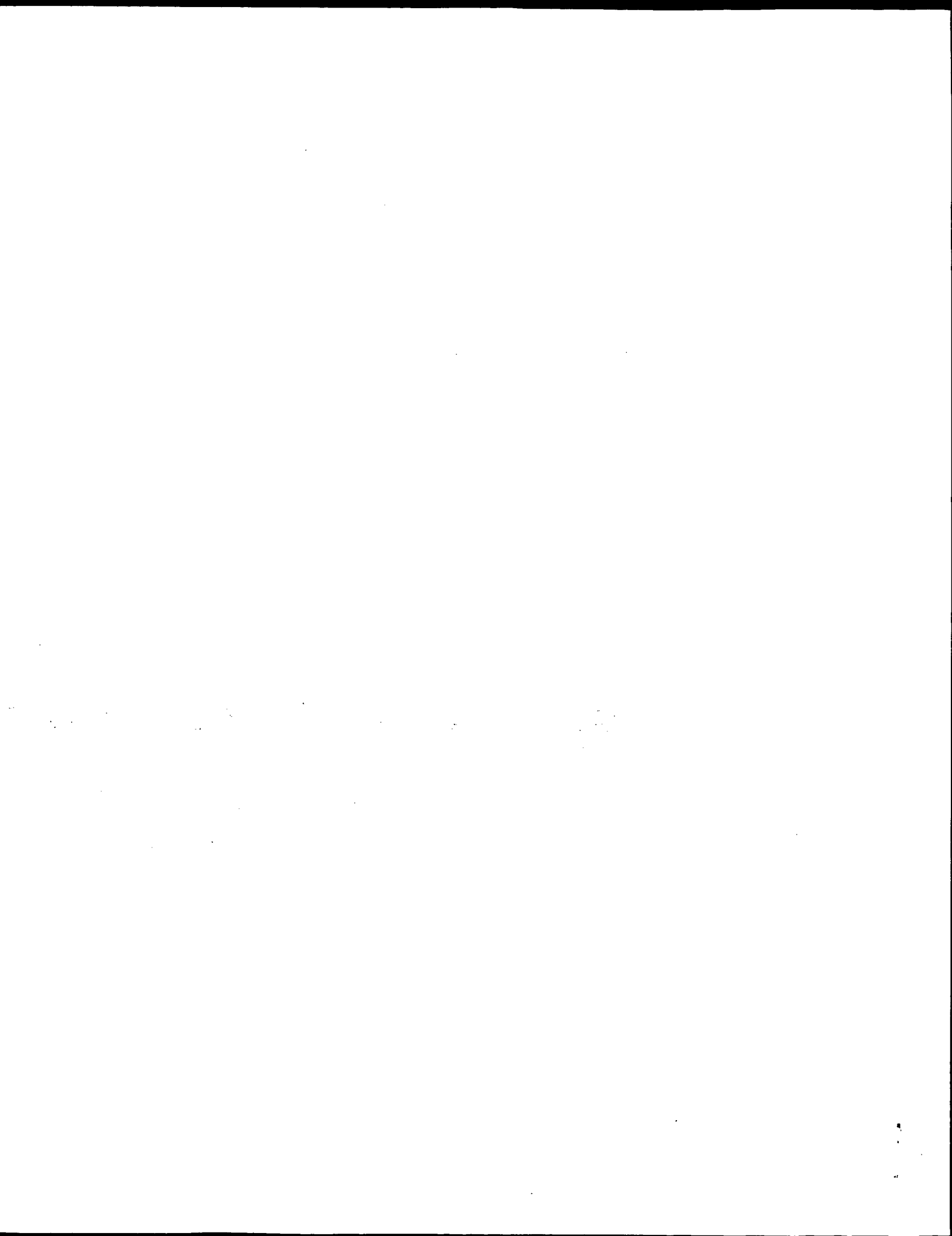
Query Match 55.8%; Score 18.4; DB 1; Length 41;  
Best Local Similarity 95.0%; Pred. No. 9.8;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 13 atggttggttacactgactg 32  
|||||  
Db 13 ATGGTTGTTACACCGACTG 32

Search completed: July 15, 2002, 22:31:51  
Job time: 18477 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:49:01 ; Search time 6165.88 Seconds

(without alignments)  
72.236 Million cell updates/sec

Title: US-10-053-641-5

Perfect score: 33

Sequence: 1 tcgggaccccttattggtgtttacactgactgc 33

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estcov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID          | Description         |
|------------|-------|-------------|--------|-------------|---------------------|
| C 1        | 23.2  | 70.3        | 954    | 12 CNS04PFY | AL301255 Tetraodon  |
| C 2        | 21.4  | 64.8        | 976    | 12 CNS002KR | AL097797 Drosophila |
| C 3        | 21    | 63.6        | 286    | 10 BF329572 | BF329572 CMO-BN027  |
| C 4        | 20.8  | 63.0        | 638    | 10 BJ069739 | BJ069739 BJ069739   |
| C 5        | 20.8  | 63.0        | 673    | 10 BJ086834 | BJ086834 BJ086834   |
| C 6        | 20.6  | 62.4        | 1090   | 12 CNS05M72 | AL343703 Tetraodon  |
| C 7        | 20.4  | 61.8        | 611    | 12 AQ271191 | AQ271191 nbxb0024K  |
| C 8        | 20.4  | 61.8        | 644    | 10 BM113581 | BM113581 EST561117  |
| C 9        | 20.4  | 61.8        | 654    | 10 BE472919 | BE472919 EST417772  |
| C 10       | 20.4  | 61.8        | 668    | 12 AG139789 | AG139789 Pan trogl  |
| C 11       | 20.4  | 61.8        | 702    | 10 BI436222 | BI436222 EST538983  |
| C 12       | 20.4  | 61.8        | 726    | 10 BI176510 | BI176510 EST517455  |
| C 13       | 20.4  | 61.8        | 739    | 12 AQ688874 | AQ688874 nbxb0078B  |
| C 14       | 20.4  | 61.8        | 859    | 10 BF030472 | BF030472 601558578  |
| C 15       | 20.4  | 61.8        | 1116   | 12 CNS06S24 | AL413894 T3 end of  |
| C 16       | 20.2  | 61.2        | 267    | 9 AW193036  | AW193036 xl69d03.x  |
| C 17       | 20.2  | 61.2        | 268    | 10 BE302044 | BE302044 bb80d03.x  |

|      |      |      |     |             |                     |
|------|------|------|-----|-------------|---------------------|
| C 18 | 20.2 | 61.2 | 314 | 9 AI377445  | AI377445 tc23f06.x  |
| C 19 | 20.2 | 61.2 | 328 | 9 AI632897  | AI632897 CX35B02.x  |
| C 20 | 20.2 | 61.2 | 367 | 10 N62121   | N62121 Yz61g08.s1   |
| C 21 | 20.2 | 61.2 | 381 | 10 N49233   | N49233 Yy83b09.s1   |
| C 22 | 20.2 | 61.2 | 456 | 10 M89275   | M89275 CEL20D3 Chr  |
| C 23 | 20.2 | 61.2 | 503 | 10 BE459104 | BE459104 EST414396  |
| C 24 | 20.2 | 61.2 | 510 | 9 AW930328  | AW930328 EST340881  |
| C 25 | 20.2 | 61.2 | 511 | 12 AZ148946 | AZ148946 SP_0003.A  |
| C 26 | 20.2 | 61.2 | 544 | 10 BG132676 | BG132676 EST465568  |
| C 27 | 20.2 | 61.2 | 565 | 10 BF098546 | BF098546 EST429067  |
| C 28 | 20.2 | 61.2 | 586 | 12 AZ176049 | AZ176049 SP_0140.B  |
| C 29 | 20.2 | 61.2 | 600 | 9 BI780646  | BI780646 EST261621  |
| C 30 | 20.2 | 61.2 | 639 | 10 BI925006 | BI925006 EST544895  |
| C 31 | 20.2 | 61.2 | 672 | 9 BB326689  | BB326689 BB326689   |
| C 32 | 20.2 | 61.2 | 912 | 12 AQ202576 | AQ202576 SP_0056.B  |
| C 33 | 20   | 60.6 | 475 | 12 AQ270392 | AQ270392 HS_2052.B  |
| C 34 | 20   | 60.6 | 707 | 12 BH016617 | BH016617 TDGCO01TH  |
| C 35 | 20   | 60.6 | 725 | 12 BH016891 | BH016891 TDGCO096TH |
| C 36 | 19.8 | 60.0 | 126 | 9 AW993381  | AW993381 RC2-BN003  |
| C 37 | 19.8 | 60.0 | 174 | 12 AZ111323 | AZ111323 RPCI-23-4  |
| C 38 | 19.8 | 60.0 | 342 | 9 BB232269  | BB232269 BB232269   |
| C 39 | 19.8 | 60.0 | 404 | 9 AI342666  | AI342666 qo31b09.x  |
| C 40 | 19.8 | 60.0 | 476 | 10 N63009   | N63009 Yy99b11.s1   |
| C 41 | 19.8 | 60.0 | 495 | 10 BG149467 | BG149467 nad29b06.  |
| C 42 | 19.8 | 60.0 | 571 | 9 AI700684  | AI700684 wa39c12.x  |
| C 43 | 19.8 | 60.0 | 608 | 12 AQ855887 | AQ855887 CpG1626B   |
| C 44 | 19.8 | 60.0 | 873 | 10 BG678841 | BG678841 602634737  |
| C 45 | 19.8 | 60.0 | 953 | 12 CNS01VMO | AL169305 Tetraodon  |

## ALIGNMENTS

### RESULT 1

CNS04PFY/c

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

CNS04PFY 954 bp DNA linear GSS 24-MAY-2000  
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone  
127M24 of library G from Tetraodon nigroviridis, genomic survey  
sequence.  
AL301255

AL301255.1 GI:8178295  
GSS: genome survey sequence.  
Tetraodon nigroviridis.  
Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 954)

Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,  
Bounau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and  
Weissenbach, J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis

Unpublished  
2 (bases 1 to 954)  
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bounau, L., Fisher, C.,  
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,  
Saurin, W. and Weissenbach, J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence

Unpublished  
3 (bases 1 to 954)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.

Location/Qualifiers  
1. .954  
/organism="Tetraodon nigroviridis"

EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 286)  
Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=CM0&t2=CM0-BN0270-  
240800-396-all&t3=2000-08-24&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 26  
High quality sequence stop: 286.  
Location/Qualifiers  
1..286  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0270"  
/dev\_stage="Adult"  
/note="Organ: breast;normal; Vector: puc18; Site:1: SmaI;  
Site:2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 80 a 58 c 73 g 75 t  
ORIGIN  
Query Match 63.6%; Score 21; DB 10; Length 286;  
Best Local Similarity 82.8%; Pred. No. 4.6e+02;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 3 ggatccttatggttggttacctgact 31  
|||||  
Db 143 GGATTTTTTATGGCTTTTACACTGACT 171  
|||||  
RESULT 4  
BJ069739/c  
LOCUS  
DEFINITION  
laevis cDNA clone XL054920 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Xenopus.  
1 (bases 1 to 638)  
Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara  
,Y.

TITLE  
JOURNAL  
COMMENT  
Expressed genes in *X. laevis* embryo  
Unpublished (2001)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers

## FEATURES

source  
1..638

/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="XL054920"  
/clone\_lib="NIBB Mochii normalized Xenopus tailbud library"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 25"

BASE COUNT 173 a 153 c 157 g 155 t  
ORIGIN

Query Match 63.0%; Score 20.8; DB 10; Length 638;  
Best Local Similarity 78.1%; Pred. NO. 5.8e+02;  
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 tcggatcctttatgtgtgtttacactgactg 32  
||||| ||| ||| ||| ||| ||| ||| |||

Db 628 TTGGGATCTTTCTGGATATTTTCACGCTG 597

## RESULT 5

## LOCUS

DEFINITION BJ086834 NIBB Mochii normalized Xenopus tailbud library Xenopus laevis cDNA clone XL054920 3', mRNA sequence.

## ACCESSION

VERSION BJ086834.1 GI:17583393

## KEYWORDS

SOURCE African clawed frog.

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.  
1 (bases 1 to 673)  
Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and Kohara, Y.

## TITLE

Unpublished (2001)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers

## FEATURES

source  
1..673

/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="XL054920"  
/clone\_lib="NIBB Mochii normalized Xenopus tailbud library"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 25"

BASE COUNT 194 a 151 c 154 g 172 t 2 others  
ORIGIN

Query Match 63.0%; Score 20.8; DB 10; Length 673;  
Best Local Similarity 78.1%; Pred. NO. 5.8e+02;  
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 tcggatcctttatgtgtgtttacactgactg 32  
||||| ||| ||| ||| ||| ||| ||| |||

Db 419 TTGGGATCTTTCTGGATATTTTCACGCTG 450

## RESULT 6

## LOCUS

DEFINITION CNS05W72 1090 bp DNA linear GSS 26-MAY-2000  
Tetraodon nigroviridis genome survey sequence T3 end of clone 039D07 of library A from Tetraodon nigroviridis, genomic survey sequence.

## ACCESSION

VERSION AL343703.1 GI:8237473

## KEYWORDS

SOURCE GSS: genome survey sequence.

## ORGANISM

Tetraodon nigroviridis.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

## REFERENCE

1 (bases 1 to 1090)  
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.

## TITLE

Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

## JOURNAL

REFERENCE Unpublished

## AUTHORS

2 (bases 1 to 1090)  
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.

## TITLE

Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

## JOURNAL

REFERENCE Unpublished

## AUTHORS

3 (bases 1 to 1090)  
Direct Submission

## TITLE

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.

## FEATURES

Location/Qualifiers

1..1090

/organism="Tetraodon nigroviridis"

/db\_xref="taxon:99883"

/clone="039D07"

/clone\_lib="A"

/note="Genoscope sequence ID : COAA039CB04A1-end : T3"

BASE COUNT 367 a 192 c 230 g 288 t 13 others  
ORIGIN

## Query Match

Best Local Similarity 62.4%; Score 20.6; DB 12; Length 1090;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 4 ggaatcctttatgtgtgtttacactgactg 32

||||| ||| ||| ||| ||| ||| ||| |||

Db 868 GGATCAATTTATGNTTGTATTACAGACAG 840

## RESULT 7

## LOCUS

DEFINITION AQ271191 611 bp DNA linear GSS 03-NOV-1998  
nxb0024K04r CUGI Rice BAC Library Oryza sativa genomic clone nxb0024K04r, DNA sequence.

## ACCESSION

VERSION AQ271191.1 GI:3824506

## KEYWORDS

SOURCE GSS.

## ORGANISM

Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;



```

Best Local Similarity 80.0%; Pred. No. 8.3e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 ggatcctttatggtttttacactgactgc 33
|||||  |||  |||  |||  |||  |||  |||  |||
Db 282 GGATCCTTTCTGTCGTGTTTACACAGACCG 311

RESULT 11
BI436222
LOCUS BI436222 702 bp mRNA linear EST 21-AUG-2001
DEFINITION BI436222 cSTE Solanum tuberosum cDNA clone cSTE22C11 5' sequence,
mRNA sequence.
ACCESSION BI436222
VERSION BI436222.1 GI:15260912
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 702)
AUTHORS van der Hoeven,R., Bezzerides,J., Bachem,C., Visser,R., Cho,J.,
Chieningo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and
Baker,B.
TITLE Generation of ESTs from in vitro grown microtubers
JOURNAL Unpublished (2001)
COMMENT Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.

FEATURES
source
location/Qualifiers
1..702
/organism="Solanum tuberosum"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="cSTE22C11"
/clone_lib="cSTE"
/tissue_type="axillary buds of stem explants; growing
sink-tubers"
/dev_stage="7, 8 and 10 days"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Tissue supplied by Christian Bachem and Richard
Visser (Department of Plant Breeding, Wageningen will
University, The Netherlands). The cSTA libraries will
attempt to capture the induction and initiation/initial
growth of the tuber in an in vitro system as described in
Bachem et al. (Plant Journal, 1996). Small microtubers
develop from axillary buds attached to stem explants when
placed on a high sucrose medium (10%). Visible
morphological changes occur synchronously at day five in
the axillary buds. The first library, cSTA (1-20) consists
of axillary buds harvested on days 1-3. This targets
those genes involved in induction of the microtubers. The
following libraries, cSTA (21-40) and cSTA (41-60),
capture genes involved in tuber initiation and outgrowth.
This library is noted as P3 in Tanksley lab notebooks."
BASE COUNT 160 a 149 c 159 g 234 t
ORIGIN

Query Match 61.8%; Score 20.4; DB 10; Length 702;
Best Local Similarity 80.0%; Pred. No. 8.3e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 ggatcctttatggtttttacactgactgc 33
|||||  |||  |||  |||  |||  |||  |||  |||
Db 552 GGAACHTTCTGTCGTGTTTACACTGCTTC 581

RESULT 12

```





# TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM498 row: 1 column: 19  
High quality sequence stop: 111.

## FEATURES source

Location/Qualifiers  
1. 859  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3828258"  
/clone\_lib="NIH\_MGC\_58"  
/tissue\_type="hypertrophoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggccgctcgcc); Site\_2: SfiI (ggccattatggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-ATCTAGAGCGCGCGGCAGATG-dT(30)BN-3',  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.35 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."  
270 a 315 c 177 g 97 t

## BASE COUNT ORIGIN

Query Match 61.8%; Score 20.4; DB 10; Length 859;  
Best Local Similarity 80.0%; Pred. No. 8.4e+02;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 gggatcctttatggtgtttacactgactg 32  
||| ||||| ||| ||| ||||| ||  
Db 78 GGGTCCTTTATTGTTATTAAACTGAATG 107

## RESULT 15 CNS06S24/c

LOCUS  
T3 end of clone AW0AA027F06 of library AW0AA from strain CLIB 89 of  
Yarrowia lipolytica, genomic survey sequence.

AL413894  
AL413894.1 GI:12186382

## ACCESSION VERSION KEYWORDS SOURCE ORGANISM

Yarrowia lipolytica.  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Dipodascaceae; Yarrowia.  
1 (bases 1 to 1116)

## REFERENCE AUTHORS

Casaregola, S., Neuvéglise, C., Lepingle, A., Bon, E., Feynerol, C.,  
Artiguenave, F., Wincker, P. and Gaillardin, C.  
Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia  
lipolytica  
FEMS Lett. 487 (1), 95-100 (2000)

## JOURNAL MEDLINE REFERENCE AUTHORS

2 (bases 1 to 1116)  
Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,  
Bollotin, F., Fukuhara, M., Bon, E., Brothier, P., Casaregola, S.,  
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Liorente, B.,  
Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S.,  
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,  
Wincker, P. and Weissenbach, J.  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies

## TITLE

# JOURNAL MEDLINE REFERENCE AUTHORS TITLE JOURNAL

FEMS Lett. 487 (1), 3-12 (2000)  
20584711  
3 (bases 1 to 1116)  
Genoscope.  
Direct Submission  
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.

## FEATURES source

Location/Qualifiers  
1. .1116  
/organism="Yarrowia lipolytica"  
/strain="CLIB 89"  
/db\_xref="taxon:4952"  
/clone="AW0AA027F06"  
/clone\_lib="AW0AA"  
/note="end : T3"

BASE COUNT 325 a 224 c 272 g 292 t 3 others  
ORIGIN

## Query Match Best Local Similarity Matches

61.8%; Score 20.4; DB 12; Length 1116;  
80.0%; Pred. No. 8.6e+02;  
24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 gggatcctttatggtgtttacactgactg 32  
||| ||||| ||||| ||||| ||||| ||  
Db 870 GGTCTCTAGATGGTGTGTACACTCATG 841

Search completed: July 15, 2002, 21:49:08  
Job time: 18899 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:29:14 ; Search time 2368.24 Seconds  
(without alignments)  
273.926 Million cell updates/sec

Title: US-10-053-641-6  
Perfect score: 31  
Sequence: 1 gctctagagccaggatgcatgtgtgccc 31

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba:\*
- 2: gb\_hg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| -----      |       |             |        |    |             |

| C | 1  | 26.2 | 84.5 | 226    | 12 | SYNHIR     |
|---|----|------|------|--------|----|------------|
| C | 2  | 26.2 | 84.5 | 235    | 6  | I41187     |
| C | 3  | 26.2 | 84.5 | 235    | 6  | I41188     |
| C | 4  | 26.2 | 84.5 | 235    | 6  | I41189     |
| C | 5  | 26.2 | 84.5 | 235    | 6  | I41190     |
| C | 6  | 25.4 | 81.9 | 59     | 6  | I63546     |
| C | 7  | 25.4 | 81.9 | 59     | 6  | I63548     |
| C | 8  | 25.4 | 81.9 | 59     | 6  | I63548     |
| C | 9  | 22.6 | 72.9 | 213245 | 10 | MMHC425018 |
| C | 10 | 22.6 | 72.9 | 273800 | 10 | AF100956   |
| C | 11 | 22.2 | 71.6 | 59     | 6  | I63542     |
| C | 12 | 22.2 | 71.0 | 90     | 6  | A18858     |
| C | 13 | 22.2 | 71.0 | 90     | 6  | A18862     |
| C | 14 | 22.2 | 71.0 | 91     | 6  | A18857     |
| C | 15 | 22.2 | 71.0 | 91     | 6  | A18861     |
| C | 16 | 22.2 | 71.0 | 195    | 6  | A03693     |
| C | 17 | 22.2 | 71.0 | 195    | 6  | A03694     |
| C | 18 | 22.2 | 71.0 | 217    | 6  | A03695     |
| C | 19 | 22.2 | 71.0 | 217    | 6  | A03696     |
| C | 20 | 22.2 | 71.0 | 217    | 6  | A34238     |
| C | 21 | 22.2 | 71.0 | 217    | 6  | A34239     |
| C | 22 | 22.2 | 71.0 | 217    | 6  | E00657     |
| C | 23 | 22.2 | 71.0 | 220    | 12 | SYNHIRMED  |
| C | 24 | 22.2 | 71.0 | 224    | 6  | A00630     |
| C | 25 | 22.2 | 71.0 | 224    | 6  | A00631     |
| C | 26 | 22.2 | 71.0 | 224    | 6  | A04277     |
| C | 27 | 22.2 | 71.0 | 224    | 6  | A04278     |
| C | 28 | 22.2 | 71.0 | 224    | 6  | A07345     |
| C | 29 | 22.2 | 71.0 | 224    | 6  | A07346     |
| C | 30 | 22.2 | 71.0 | 279    | 6  | A01139     |
| C | 31 | 22.2 | 71.0 | 279    | 6  | I26634     |
| C | 32 | 21.8 | 70.3 | 32     | 6  | A00638     |
| C | 33 | 21.8 | 70.3 | 64     | 6  | A34232     |
| C | 34 | 21.8 | 70.3 | 109    | 6  | A34234     |
| C | 35 | 21.8 | 70.3 | 109    | 6  | A34235     |
| C | 36 | 21.4 | 69.0 | 39840  | 2  | AC016877   |
| C | 37 | 21.4 | 69.0 | 157952 | 2  | AP003691   |
| C | 38 | 21.4 | 69.0 | 160952 | 9  | AP003551   |
| C | 39 | 21.2 | 68.4 | 136613 | 2  | AC090379   |
| C | 40 | 21.2 | 67.7 | 118121 | 2  | AC034189   |
| C | 41 | 21.2 | 67.7 | 149466 | 9  | AC080100   |
| C | 42 | 21.2 | 67.7 | 173169 | 2  | AC094464   |
| C | 43 | 21.2 | 67.7 | 179438 | 9  | AC026866   |
| C | 44 | 21.2 | 67.7 | 194547 | 9  | AC018951   |
| C | 45 | 20.8 | 67.1 | 40     | 6  | A18123     |

## ALIGNMENTS

| RESULT     | 1 | SYNHIR/c | SYNHIR  | 226 bp | DNA | linear | SYN 27-APR-1993 |
|------------|---|----------|---|--------|-----|--------|-----------------|
| LOCUS      |   |          | Synthetic hirudin gene, complete cds.   |        |     |        |                 |
| DEFINITION |   |          | Synthetic hirudin gene, complete cds.   |        |     |        |                 |
| ACCESSION  |   |          | M26762  |        |     |        |                 |
| VERSION    |   |          | M26762.1  |        |     |        |                 |
| KEYWORDS   |   |          | hirudin; proteinase inhibitor; thrombin inhibitor.  |        |     |        |                 |
| SOURCE     |   |          | Synthetic DNA.  |        |     |        |                 |
| ORGANISM   |   |          | synthetic construct   |        |     |        |                 |
| REFERENCE  |   |          | 1 (bases 1 to 226)  |        |     |        |                 |
| AUTHORS    |   |          | Bergmann,C., Drott,J., Koehler,S., Fink,E. and Gassen,H.G.  |        |     |        |                 |
| TITLE      |   |          | Chemical synthesis and expression of a gene coding for hirudin, the thrombin-specific inhibitor from the leech Hirudo medicinalis |        |     |        |                 |
| JOURNAL    |   |          | Biol. Chem. Hoppe-Seyler 367, 731-740 (1986)  |        |     |        |                 |
| MEDLINE    |   |          | 87026239  |        |     |        |                 |
| FEATURES   |   |          | Location/Qualifiers   |        |     |        |                 |
|            |   |          | 1..226  |        |     |        |                 |
|            |   |          | /organism="synthetic construct"   |        |     |        |                 |
|            |   |          | /db_xref="taxon:32630"  |        |     |        |                 |
|            |   |          | 22..222   |        |     |        |                 |
|            |   |          | /note="hirudin"   |        |     |        |                 |
|            |   |          | /codon_start=1  |        |     |        |                 |
|            |   |          | CDS   |        |     |        |                 |



```

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 59)
AUTHORS Kurfuerst,M., Ruebsamen,K., Schmied,B., Koerwer,W., Schweden,J. and
         Hoeffken,H.Wolfgang.
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins
JOURNAL Patent: US 5663141-A 9 02-SEP-1997;
FEATURES Location/Qualifiers
          source
            1..59
BASE COUNT 20 a 14 c 15 g 10 t
ORIGIN

Query Match
Best Local Similarity 81.9%; Score 25.4; DB 6; Length 59;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gctctagagcccgagtcattgtt 27
Db 27 GCCTTTAGAGCCGAGTCATTGTT 1

RESULT 7
I63548/c
LOCUS I63548
DEFINITION Sequence 11 from patent US 5663141.
ACCESSION I63548
VERSION I63548.1 GI:2481121
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 59)
AUTHORS Kurfuerst,M., Ruebsamen,K., Schmied,B., Koerwer,W., Schweden,J. and
         Hoeffken,H.Wolfgang.
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins
JOURNAL Patent: US 5663141-A 11 02-SEP-1997;
FEATURES Location/Qualifiers
          source
            1..59
BASE COUNT 18 a 14 c 16 g 11 t
ORIGIN

Query Match
Best Local Similarity 81.9%; Score 25.4; DB 6; Length 59;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gctctagagcccgagtcattgtt 27
Db 27 GCCTTTAGAGCCGAGTCATTGTT 1

RESULT 8
I41183/c
LOCUS I41183
DEFINITION Sequence 3 from patent US 5624822.
ACCESSION I41183
VERSION I41183.1 GI:2081773
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 129)
AUTHORS Koerwer,W.
TITLE Hirudin fusion proteins and preparation of hirudin
JOURNAL Patent: US 5624822-A 3 29-APR-1997;
FEATURES Location/Qualifiers
          source
            1..129
BASE COUNT 36 a 35 c 29 g 29 t
ORIGIN

```

```

Query Match
Best Local Similarity 80.6%; Score 25; DB 6; Length 129;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 agagcccgagtcattgttgcgc 31
Db 129 AGAGCCCGAGTCATTGTGCCC 105

RESULT 9
MMHC425018
LOCUS MMHC425018
DEFINITION Mus musculus major histocompatibility complex region NG27, NG28,
RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein,
BING1, Rafkin, RalGDS-like, KE2, BING4, beta 1.3-galactosyl
transferase, and RPS18 genes, complete cds; Sacm21 gene, partial
cds; and unknown gene.
ACCESSION AF110520
VERSION AF110520.1 GI:4050090
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 213245)
AUTHORS Rowen,L., Qin,S., Madan,A., Loretz,C., Hall,J., James,R., Dors,M.,
Shaffer,T., Abbasi,N., Ratcliffe,A., Dickhoff,R., Lasky,S. and
Hood,L.
TITLE Sequence of the mouse major histocompatibility complex class II
region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 213245)
AUTHORS Rowen,L.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1998) Department of Molecular Biotechnology, Box
357730 University of Washington, Seattle, Washington 98195, USA
COMMENT Sequencing methodology: high redundancy shotgun. Interspersed
Repeats were identified with RepeatMasker (available from
http://ftp.genome.washington.edu/RM/RepeatMasker.html). This clone
overlaps clone 445d10, found in GenBankAccession Number AF100956.
It is possible that this clone is aberrant, since it does not
contain the SYNGAP ATPase found 5' of HSET in the human counterpart
region (kinesin, in this sequence). It is not known yet whether
this clone has deleted the SYNGAP gene or whether the 129 strain
has deleted it.
FEATURES Location/Qualifiers
          source
            1..213245
              /organism="Mus musculus"
              /strain="129"
              /db_xref="taxon:10090"
              /chromosome="17"
              /clone="BAC425018"
              /clone_lib="Genome Systems"
              /complement(422..504)
              /rpt_family="PB1D10"
              /complement(1421..1479)
              /rpt_family="PB1D10"
              /complement(1421..1553)
              /rpt_family="MMB1F"
              /complement(1514..1642)
              /rpt_family="B4A"
              /complement(1568..1642)
              /rpt_family="MUSID5"
              /complement(1895..2375)
              /rpt_family="LINE2"
              /complement(2573..2713)
              /rpt_family="MMB1F"
              /complement(2719..2793)
              /rpt_family="MUSID4"
              /complement(2951..3048)
              /rpt_family="PB1D10"

```

```

repeat_region complement(3157..3276)
repeat_region /rpt_unit=GA
repeat_region complement(3328..3359)
repeat_region /rpt_family="MIR"
repeat_region 3890..3962
repeat_region /rpt_unit=TAAA
repeat_region complement(5643..5705)
repeat_region /rpt_family="PB1D10"
repeat_region 7199..7335
repeat_region /rpt_family="MMB1"
repeat_region complement(7390..7480)
repeat_region /rpt_family="MMB1F"
repeat_region complement(7538..7645)
repeat_region /rpt_family="MMSQR2"
repeat_region complement(7674..8071)
repeat_region /rpt_family="MTA"
repeat_region complement(8071..8121)
repeat_region /rpt_family="PB1D10"
repeat_region complement(8246..8349)
repeat_region /rpt_family="PB1D10"
repeat_region 8717..8756
repeat_region /rpt_family="GC-rich"
mRNA complement(<8978..>9017)
CDS /product="unknown"
complement(<8978..9017)
/notes="unknown: Similar to ESTs AA260075 and AA144156.
Similar to an exon in cDNAs in L26528 and D01046"
/codon_start=1
/protein_id="AAC97964.1"
/db_xref="GI:4050091"
/translation="MGTRDDYDLFK"
repeat_region complement(9622..9702)
repeat_region /rpt_unit=GA
repeat_region complement(9775..9811)
repeat_region /rpt_family="POLY_A"
repeat_region complement(9812..9950)
repeat_region /rpt_family="MMB1F"
repeat_region complement(9965..10006)
repeat_region /rpt_unit=CA
repeat_region 1047..10159
repeat_region /rpt_unit=GA
repeat_region 10434..10612
repeat_region /rpt_family="MMB2"
repeat_region complement(10753..10952)
repeat_region /rpt_family="B3"
repeat_region 11202..11330
repeat_region /rpt_family="B3"
repeat_region complement(11348..11433)
repeat_region /rpt_unit=GA
repeat_region complement(11468..11605)
repeat_region /rpt_family="MMB1"
repeat_region complement(11627..11770)
repeat_region /rpt_family="MMB1F"
repeat_region complement(11909..12232)
repeat_region /rpt_family="ORR1A1"
repeat_region 12413..12472
repeat_region /rpt_family="MMB1F"
repeat_region 13085..13229
repeat_region /rpt_family="MTD"
repeat_region complement(13321..13598)
repeat_region /rpt_family="MMETNLTR"
repeat_region 13597..13778
repeat_region /rpt_family="MTD"
repeat_region complement(13948..14046)
repeat_region /rpt_family="MMB1F"
repeat_region 14293..14381
repeat_region /rpt_family="PB1"
repeat_region 14442..14558
repeat_region /rpt_family="MMB1"
repeat_region 14598..14673
repeat_region /rpt_family="B4"
repeat_region 14629..14724
repeat_region /rpt_family="MMSQR2"

```

```

repeat_region 14851..14991
repeat_region /rpt_family="ORR1B-INT"
repeat_region 14880..15016
repeat_region /rpt_family="MMB1"
repeat_region 15781..15911
repeat_region /rpt_family="MMB1F"
repeat_region complement(15997..16184)
repeat_region /rpt_family="B3"
repeat_region 16299..16422
repeat_region /rpt_family="MMB1F"
repeat_region 16429..16463
repeat_region /rpt_unit=CA
repeat_region 16479..16707
repeat_region /rpt_family="MMB2"
repeat_region 16719..16923
repeat_region /rpt_family="B3"
repeat_region complement(17436..17469)
repeat_region /rpt_family="AT-rich"
repeat_region complement(17472..17515)
repeat_region /rpt_unit=CA
repeat_region complement(17558..17625)
repeat_region /rpt_family="B4"
repeat_region 17707..17824
repeat_region /rpt_family="PB1D9"
repeat_region complement(17892..17976)
repeat_region /rpt_family="MMSQR2"
repeat_region complement(18032..18164)
repeat_region /rpt_unit=GA
repeat_region complement(18165..18233)
repeat_region /rpt_family="MMB2"
repeat_region complement(18229..18274)
repeat_region /rpt_family="B3"
repeat_region complement(18327..18499)
repeat_region /rpt_family="MMB2"
repeat_region 18570..18781
repeat_region /rpt_family="B4"
repeat_region 18656..18781
repeat_region /rpt_family="PB1D10"
repeat_region complement(18791..18827)
repeat_region /rpt_unit=GA
repeat_region complement(18828..18972)
repeat_region /rpt_family="MMB1"
repeat_region 19004..19085
repeat_region /rpt_family="B4A"
repeat_region 19097..19266
repeat_region /rpt_family="MMB2"
repeat_region 19303..19402
repeat_region /rpt_family="B4A"
repeat_region 19402..19551
repeat_region /rpt_family="MMB1"
repeat_region 19556..19594
repeat_region /rpt_unit=TAAA
repeat_region complement(20527..20633)
repeat_region /rpt_family="MMB1"
repeat_region complement(20740..20769)
repeat_region /rpt_unit=GA
repeat_region complement(20810..20831)
repeat_region /rpt_family="AT-rich"
repeat_region complement(20846..20945)
repeat_region /rpt_family="B3"
repeat_region complement(21057..21094)
repeat_region /rpt_family="MUSID5"
repeat_region complement(21115..21239)
repeat_region /rpt_family="PB1D7"
repeat_region 21474..21553
repeat_region /rpt_family="B3"
repeat_region 21570..21656

```

```

Query Match 72.9%; Score 22.6; DB 10; Length 213245;
Best Local Similarity 86.2%; Pred. No. 4.2;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 cctctagaccagatgcattgttgc 30

```

```

|||||  ||||||| ||||||| |||
Db 166306 CCTCTACCTCCAGGATGATTTGGCC 166334

RESULT 10
AF100956
LOCUS
DEFINITION
Mus musculus major histocompatibility locus class II region;
Fas-binding protein Daxx (DAXX) gene, partial cds; Bing1 (BING1),
tapasin (tapasin), RalGDS-like factor (RLF), KE2 (KE2), BING4
(BING4), betal, 3-galactosyl transferase (betal, 3-galactosyl
transferase), ribosomal protein subunit S18 (RPS18), Sacm21
(Sacm21), H2K1(b) (H2-K1(b)), RING1 (RING1), KE6a (KE6a), KE4
(KE4), RXRbeta (RXRbeta), collagen alpha-2 (X1) (COLA11A2), H2-O
alpha (H2-Oalpha), RING3 (RING3), H2-M alpha (H2-M alpha), H2-M
beta 2 (H2-M beta2), and H2-M beta1 (H2-M beta1) genes, complete
cds; and LMP 2 gene, partial cds.

ACCESSION
AF100956
VERSION
AF100956.1
KEYWORDS
GI:3811374
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 273800)
Rowen,L., Qin,S., Madan,A., Loretz,C., James,R., Dors,M., Mix,L.,
Hall,J., Lasky,S. and Hood,L.
AUTHORS
Sequence of the mouse major histocompatibility locus class II
region
TITLE
Unpublished
JOURNAL
2 (bases 1 to 273800)
Rowen,L.
AUTHORS
Direct Submission
JOURNAL
Submitted (26-OCT-1998) Department of Molecular Biotechnology, Box
357730 University of Washington, Seattle, WA 98195, USA
COMMENT
This sequence overlaps the entry in GenBank Accession Number
AF027865 by 22022 bases. Sequencing methodology: high redundancy
shotgun, using M13 templates. Interspersed Repeats were identified
with RepeatMasker (available from
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES
Location/Qualifiers
1..273800
/organism="Mus musculus"
/strain="129svj"
/db_xref="taxon:10090"
/chromosome="17"
1..139758
/organism="Mus musculus"
/strain="129svj"
/db_xref="taxon:10090"
/clone="BAC 445D10"
/clone_lib="Genome Systems"
<1..2000
/gene="DAXX"
Join(<146..208,325..544,1825..2000)
/gene="DAXX"
/product="Fas-binding protein Daxx"
Join(<146..208,325..544,1825..1884)
/gene="DAXX"
/note="intron-exon boundaries defined in relation to cDNA
sequence in GenBank Accession Number AF006040"
/codon_start=2
/product="Fas-binding protein Daxx"
/protein_id="AAC69891.1"
/db_xref="GI:3811375"
/translation="PPSKRFRKKEKKQLGSLGNGSYIKEPMAQODSQNTSVQPMPSPP
PLASVASVDSSSTRVDSPSHLVTSLSLSPSPSLLOTPQAQSLRQCIYKTSVATQCD
PEEIVLSDSQ"
724..858
/rpt_family="MMB1"
complement(1257..1297)
/rpt_family="(CA)n"
1303..1421
repeat_region
repeat_region
repeat_region

```

```

/rpt_family="MMB1F"
complement(1426..1630)
/rpt_family="B3"
complement(1646..1792)
/rpt_family="MMB1F"
complement(2728..2788)
/rpt_family="GC-rich"
3168..3214
/rpt_family="MIR"
<3365..>5281
/gene="BING1"
/product="Bing1"
<3365..>5281
/gene="BING1"
/note="Zn finger; mouse orthologue to BING 1, described in
GenBank Accession Number 297183"
3365..5281
/gene="BING1"
/note="this orf was predicted by Genscan and confirmed by
ESTs AA538332 and AA797063"
/codon_start=1
/product="Bing1"
/protein_id="AAC69892.1"
/db_xref="GI:3811376"
/translation="MEPSALSPSGATLPLPLSLAPPLPLPAAAVVHVSPFVTSALL
ESLNQRLQGLCDVSIQVGRFRAHRAVLAASSPYFHDVLLKMGTSISLPSVMDP
GAFETVLASAYTGLSMAADIIVFELTVGSLQMHIVDKTELLREGSAAATTTVT
TAAAPSVSPCASVPNGGTVPATVSVRSHTSRASSENSSPSSNFPSPSTDF
SSYQDAFVASAAGSGNRDGGVPFAPVVGVSAGTSGKLLLELDCDDGGDGRGAV
APGALRRNCMPASAPQKHVVYVQARNCPASLVHQDLEDEEEDLVLTCE
DDEDEMGSGSVGAGEPEATLSIDVRTLTPADKGEQNFCESSNDPFGYEGGG
AGALDDPGGPTPSSYALTPRPRLPLDVPNGNOLVFPSSSOAQPDPGNTAEHGA
VTLGSTVAVGLGIPSGSGGAPGTGSDGNKIFLCHGKAFSHKSRDRHVNHLNR
PFCPCVNCNKKFKMKHHLTTHMKTHTGLKPYECVCAKKEKFMKSDSPWRHRCCHERRHM
GVGVGSGGPGPGPGTGPQLPKRESYTGSGVGGDEANSATPPSHRRVWSPSPSVH
KVMDFSGGGGAH"
Join(6121..6206,6353..6532,6731..6991,11888..12286,
12486..12827,13108..13184,14618..14754)
/gene="tapasin"
/product="tapasin"
<6121..>14754
/gene="tapasin"
Join(6170..6206,6353..6532,6731..6991,11888..12286,
12486..12827,13108..13184,14618..14629)
/gene="tapasin"
/note="intron-exon boundaries defined in relation to cDNA
in GenBank Accession Numbers AF043943 and EST AA362898;
EST W75528 contains an alternatively spliced form of this
transcript"
/codon_start=1
/product="tapasin"
/protein_id="AAC69893.1"
/db_xref="GI:3811377"
/translation="MKPLLLVAVALLGLATVYVSVSAGPEAIECFWEDAGGGGLSKK
PATLLRHGRPGPPPCRALDGLKLYFKVDDPAGMLLAFFRRYPAGASAPHCMSRFTPF
PASAKWASLSLSPENCPRALDGLVSVSTLSLSLLRPQEPLEPVPVITWATV
VLTVLTHNPAPRVQLGKDAVLDLRFAYAPSALEGSPLDAGPPPGLEWRGRKNGH
LLIAATPLAGLRMPPAQAKATATAFADDDDEPGWGTGNGTIFWLPAPVKPQEGVLYLATV
HIPYLOGVSLTLTHVKARATSLTPAPVYVAAAPGAEPALCLASHFFPAEGLEVYKWE
LRGPGGSRKRVKGTWLTSTIRHSDGSYSQSHGLQLPPTATKQHGQVHYVCRVYHSSL
PASGRADVTLEVAAYWTIPEVSKKATAASLTIPRNSKKSQ"
complement(8176..8383)
/rpt_family="B3"
8345..8670
/rpt_family="B3"
complement(9036..9243)
/rpt_family="MMB2"
complement(9853..9997)
/rpt_family="MMB1"
10003..10014
/gene="tapasin"
/note="number of Ts unclear"
complement(10003..10198)
repeat_region

```









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:53 ; Search time 758.37 Seconds  
(without alignments)  
70.183 Million cell updates/sec

Title: US-10-053-641-6  
Perfect score: 31  
Sequence: 1 gctctagaccagatgcattgttgc 31

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N.Geneseq\_032802.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| C 1        | 29.4  | 94.8        | 468    | 16 | AAQ93244    |
| C 2        | 26.2  | 84.5        | 726    | 12 | AAQ12379    |
| C 3        | 22    | 71.0        | 183    | 12 | AAQ14927    |
| C 4        | 22    | 71.0        | 185    | 12 | AAQ14926    |
| C 5        | 22    | 71.0        | 196    | 8  | AAQ70319    |
| C 6        | 22    | 71.0        | 217    | 7  | AAQ60355    |
| C 7        | 22    | 71.0        | 217    | 8  | AAQ70323    |
| C 8        | 22    | 71.0        | 224    | 7  | AAQ60746    |
| C 9        | 22    | 71.0        | 279    | 12 | AAQ13831    |
| C 1        | 29.4  | 94.8        | 468    | 16 | AAQ93244    |
| C 2        | 26.2  | 84.5        | 726    | 12 | AAQ12379    |
| C 3        | 22    | 71.0        | 183    | 12 | AAQ14927    |
| C 4        | 22    | 71.0        | 185    | 12 | AAQ14926    |
| C 5        | 22    | 71.0        | 196    | 8  | AAQ70319    |
| C 6        | 22    | 71.0        | 217    | 7  | AAQ60355    |
| C 7        | 22    | 71.0        | 217    | 8  | AAQ70323    |
| C 8        | 22    | 71.0        | 224    | 7  | AAQ60746    |
| C 9        | 22    | 71.0        | 279    | 12 | AAQ13831    |

|      |      |      |        |    |           |                     |
|------|------|------|--------|----|-----------|---------------------|
| C 10 | 20.8 | 67.1 | 201    | 12 | AAQ12153  | Synthetic hirudin   |
| C 11 | 20.8 | 67.1 | 201    | 13 | AAQ25184  | Hirudin HV-1. Syn   |
| C 12 | 20.8 | 67.1 | 420    | 12 | AAQ12155  | Factor Xa-cleavabl  |
| C 13 | 20.8 | 67.1 | 1458   | 12 | AAQ12162  | Factor Xa-cleavabl  |
| C 14 | 20.8 | 67.1 | 1467   | 12 | AAQ12490  | Factor Xa-cleavabl  |
| C 15 | 20.4 | 65.8 | 2300   | 24 | AAQ62246  | CDNA sequence #33   |
| C 16 | 20.4 | 65.8 | 2864   | 22 | AAH76458  | CDNA corresponding  |
| C 17 | 20.2 | 65.2 | 1618   | 21 | AAQ44088  | zee may's DNA fragm |
| C 18 | 20   | 64.5 | 9048   | 21 | AAA64507  | Nucleotide sequenc  |
| C 19 | 20   | 64.5 | 10556  | 22 | AAI59459  | Human poly nucleoti |
| C 20 | 19.8 | 63.9 | 828    | 13 | AAQ24269  | Encodes PLGF angio  |
| C 21 | 19.8 | 63.9 | 3543   | 22 | AAH18248  | Human CDNA sequenc  |
| C 22 | 19.6 | 63.2 | 104    | 21 | AAQ13076  | Human secreted pro  |
| C 23 | 19.2 | 61.9 | 227    | 22 | AAQ61507  | S. marcescens hiru  |
| C 24 | 19.2 | 61.9 | 229    | 9  | AAQ81294  | Sequence encoding   |
| C 25 | 19.2 | 61.9 | 238    | 9  | AAQ81295  | Sequence encoding   |
| C 26 | 19.2 | 61.9 | 1458   | 22 | AAH68269  | C glutamicum codin  |
| C 27 | 19.2 | 61.9 | 309400 | 22 | AAH68534  | C glutamicum codin  |
| C 28 | 19   | 61.3 | 969    | 23 | ABL25309  | Drosophila melanog  |
| C 29 | 19   | 61.3 | 1817   | 22 | AAQ21961  | Human collagen gen  |
| C 30 | 19   | 61.3 | 2906   | 23 | ABL25306  | Drosophila melanog  |
| C 31 | 19   | 61.3 | 2969   | 23 | ABL25308  | Drosophila melanog  |
| C 32 | 19   | 61.3 | 10234  | 22 | AAQ105428 | Human reproductive  |
| C 33 | 19   | 61.3 | 20084  | 22 | AAQ21773  | Human gene for col  |
| C 34 | 19   | 61.3 | 20084  | 22 | AAQ83656  | Nucleotide sequenc  |
| C 35 | 19   | 61.3 | 35959  | 22 | AAQ78275  | Human immune/haema  |
| C 36 | 19   | 61.3 | 50885  | 22 | AAQ70336  | Human immune/haema  |
| C 37 | 18.8 | 60.6 | 198    | 13 | AAQ24986  | Sequence encoding   |
| C 38 | 18.8 | 60.6 | 210    | 12 | AAQ10452  | HV-1 gene. Synthe   |
| C 39 | 18.8 | 60.6 | 430    | 22 | AAQ125702 | Human breast cance  |
| C 40 | 18.8 | 60.6 | 453    | 22 | AAQ18097  | Human breast cance  |
| C 41 | 18.8 | 60.6 | 456    | 22 | AAQ17956  | Human breast cance  |
| C 42 | 18.8 | 60.6 | 552    | 23 | ABL10237  | Drosophila melanog  |
| C 43 | 18.8 | 60.6 | 665    | 21 | AAQ43982  | Mouse secreted exp  |
| C 44 | 18.8 | 60.6 | 584    | 19 | AAQ20815  | Rat haematopoietic  |
| C 45 | 18.8 | 60.6 | 1360   | 24 | ABQ199523 | Mouse ischaemic co  |

ALIGNMENTS

RESULT 1  
AAQ93244/c  
ID AAQ93244 standard; DNA; 468 BP.  
XX AC AAQ93244;  
XX DT 01-NOV-1995 (first entry)  
XX DE Fusion construct of glucoamylase-hirudin DNA.  
XX KW expression cassette; recombinant protein; production; Hansenula;  
KW yeast; leader sequence; adaptor; alpha helix; glucoamylase;  
KW secretion; processing; thrombin inhibitor; hirudin; ds.  
XX OS Synthetic.  
XX FH Key  
FH primer\_bind Location/Qualifiers  
FT complement (1..33)  
FT /tag= a  
FT /note= "primer AAQ85840 binding site"  
FT CDS 13..444  
FT /tag= b  
FT /product= glucoamylase-hirudin fusion protein  
FT primer\_bind 199..223  
FT /tag= c  
FT /note= "primer AAQ85841 binding site"  
FT primer\_bind complement (220..267)  
FT /tag= d  
FT /note= "primer AAQ85842 binding site"  
FT misc\_feature 13..228  
FT /tag= e  
FT /note= "encodes amino acids 1-72 of glucoamylase"

FT misc\_feature 241..246  
 FT /\*tag= f  
 FT /note= "encodes processor signal"  
 FT misc\_feature 247..444  
 FT /\*tag= g  
 FT /note= "encodes hirudin-h120"  
 FT primer\_bind 0..468  
 FT /\*tag= h  
 FT /note= "primer AAQ85843 binding site"  
 XX  
 PN DE4329969-A.  
 XX  
 XX 09-MAR-1995.  
 XX  
 XX 04-SEP-1993; 93DE-4329969.  
 XX  
 XX 04-SEP-1993; 93DE-4329969.  
 XX  
 XX (BADI ) BASF AG.  
 PA (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.  
 XX  
 XX Bollschweiler C, Janowicz ZA, Piontek M, Schweden J;  
 PI Strasserawm, Weydemann U;  
 XX  
 DR WPI; 1995-107810/15.  
 DR P-PSDB; AAR76951.  
 XX

PT Recombinant protein prodn. in Hansenula yeast - transformed with  
 PT expression cassette contg. leader, adaptor, processing signal and  
 PT gene, provides efficient secretion and correct processing  
 XX  
 PS Example 1; Fig 1; 10pp; German.  
 XX  
 CC AAQ93244 is a DNA fusion product resulting from ligation of an  
 CC EcoRI-PvuI fragment (see AAQ85840-41) and a PvuI-SalI fragment, encoding  
 CC a leader sequence, an adaptor (see AAR71472, amino acids 23-72 of GAM  
 CC (glucanylase from Schwannomyces occidentalis, plus -His-Pro-Leu-cln at  
 CC the C-terminal end), a processor peptide (Lys-Arg) and a structural gene,  
 CC encoding hirudin. AAQ93244 is an example of an expression vector insert  
 CC of the general formula: L-A-P-Gene (L = leader sequence; A = adaptor;  
 CC P = processor; Gene = structural gene). The cassettes ensure efficient  
 CC secretion and correct processing of heterologous structural genes in  
 CC yeast of the genus Hansenula, and so provides high yields of mature  
 CC proteins and facilitates subsequent purification.  
 XX  
 SQ Sequence 468 BP; 126 A; 110 C; 112 G; 120 T; 0 other;

Query Match 94.8%; Score 29.4; DB 16; Length 468;  
 Best Local Similarity 96.8%; Pred. No. 0.0015;  
 Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 gcctctagagccagcagcattgttgcgc 31  
 ||| |||||  
 DB 348 GCCTTTAGAGCCAGGATGCTTTGTTGCC 318

RESULT 2  
 AAQ12379/C  
 ID AAQ12379 standard; DNA; 726 BP.  
 XX  
 AC AAQ12379;  
 XX  
 XX 17-SEP-1991 (first entry)  
 XX  
 DE Hirudin peptide/Protein A fusion gene.  
 XX  
 KW anticoagulant; fusion protein; ds.  
 XX  
 XX Key Location/Qualifiers  
 FT mat\_peptide 1..489  
 FT /\*tag= a  
 FT /product= Protein A

FT mat\_peptide 529..720  
 FT /\*tag= b  
 FT /product= Hirudin  
 FT misc\_RNA 490..528  
 FT /\*tag= c  
 FT /product= linker oligopeptide Y

FT DE3942580-A.  
 PN 27-JUN-1991.  
 XX  
 XX 22-DEC-1989; 89DE-3942580.  
 PF  
 XX 22-DEC-1989; 89DE-3942580.  
 PR  
 XX (BADI ) BASF AG.  
 PA  
 XX Korwer W;  
 PI  
 XX WPI; 1991-1942336/27.  
 DR P-PSDB; AAR12751.  
 XX

XX Hirudin peptide prodn. by cleaving new fusion peptide - of  
 PT hirudin and protein A, expressed in high yield and stable,  
 PT soluble form by transformed E.coli  
 PT  
 XX Example 1; Page 6-7; 9pp; German.

XX This sequence is an example of a fusion construct for expression of  
 CC the fusion peptide of the invention. The Met residue in the linker  
 CC oligopeptide allows cleavage by CNBr to release two fragments which  
 CC can be easily separated by IgG affinity chromatography. (The protein  
 CC A component binds to IgG sepharose). Increased yields of hirudin are  
 CC obtained using this fusion construct.  
 XX  
 SQ Sequence 726 BP; 280 A; 171 C; 126 G; 149 T; 0 other;

Query Match 84.5%; Score 26.2; DB 12; Length 726;  
 Best Local Similarity 90.3%; Pred. No. 0.045;  
 Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 gcctctagagccagcagcattgttgcgc 31  
 ||| |||||  
 DB 627 GCCTCAGAGCCAGGATGCTTTGTTGCC 597

RESULT 3  
 AAQ14927/C  
 ID AAQ14927 standard; DNA; 183 BP.  
 XX  
 AC AAQ14927;  
 XX  
 XX 24-FEB-1992 (first entry)  
 DT  
 DE Partial Hirudin HV12 coding sequence.  
 XX  
 XX anti-thrombin activity; thrombosis; HV1; HV2; blood clotting; ds.  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 FT CDS 7..177  
 FT /\*tag= a  
 FT /note= "encodes residues 10-65 of HV12"

XX WO9117250-A.  
 PN  
 XX 14-NOV-1991.  
 PD  
 XX 05-APR-1991; 91WO-BP00643.  
 PF  
 XX 10-MAY-1990; 90GB-0010552.  
 PR

XX (FARM ) FARMITALIA C ERBA SRL.  
 XX Benatti L, Carminati P, Lansen J, Mazue G, Roncucci R;  
 PI Sarmientosp, Scacheri E, De Taxis du Poet P;  
 XX WPI; 1991-353771/48.  
 XX  
 XX Expression vectors encoding new and known hirudin(s) - or  
 PT hirudin-like polypeptide(s), useful for treating thromboembolic  
 PT events and thromboses etc.  
 XX  
 XX Example 1; Fig 3; 41pp; English.  
 XX  
 CC Hybrid hirudin HV12 is a HVI derivative composed of the first 46  
 CC residues of HVI followed by the amino acid sequence from residue  
 CC 47-65 of the HV2 variant. A synthetic gene encoding the new deriv.  
 CC was designed based on the preferred E.coli codons. The  
 CC double-stranded construct, assembled from four oligonucleotides,  
 CC has HindIII and PstI sticky ends. See AAR15147 for HV12 sequence.  
 XX  
 XX Sequence 183 BP; 55 A; 40 C; 47 G; 41 T; 0 other;  
 SQ

Query Match 71.0%; Score 22; DB 12; Length 183;  
 Best Local Similarity 83.3%; Pred. No. 2.6;  
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cctctagagccagcagtcattgttgc 31  
 II III IIIIIIIIIIIIIIIII III  
 DB 80 CCGTCAGAACCCAGGATGTCATTGTACCC 51

RESULT 4  
 AAQ14926/c  
 ID AAQ14926 standard; DNA; 185 BP.  
 XX  
 AC AAQ14926;  
 XX  
 DT 24-FEB-1992 (first entry)  
 XX  
 DE Synthetic hirudin HVI-encoding sequence.  
 XX  
 KW anti-thrombin activity; thrombosis; blood clotting; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN W09117250-A.  
 XX  
 PD 14-NOV-1991.  
 XX  
 PF 05-APR-1991; 91WO-EP00643.  
 XX  
 PR 10-MAY-1990; 90GB-0010552.  
 XX  
 XX (FARM ) FARMITALIA C ERBA SRL.  
 XX Benatti L, Carminati P, Lansen J, Mazue G, Roncucci R;  
 PI Sarmientosp, Scacheri E, De Taxis du Poet P;  
 XX WPI; 1991-353771/48.  
 XX  
 XX Expression vectors encoding new and known hirudin(s) - or  
 PT hirudin-like polypeptide(s), useful for treating thromboembolic  
 PT events and thromboses etc.  
 XX  
 XX Example 1; Fig 1; 41pp; English.  
 XX  
 CC The coding sequence was designed on the basis of the E.coli  
 CC preferred codons. A Bali restriction site was engineered close to  
 CC the 5' end of the synthetic gene to allow insertion of the coding  
 CC sequence in different expression vectors. The same synthetic gene  
 CC can be used for expression in both bacterial and insect cells (for

CC insect cells methods were developed which yielded secreted or  
 CC cytoplasmic products. The HVI gene was constructed from four  
 CC oligonucleotides and the double-stranded sequence has HindII and  
 CC PstI sticky ends.  
 XX  
 XX Sequence 185 BP; 52 A; 44 C; 51 G; 38 T; 0 other;  
 SQ

Query Match 71.0%; Score 22; DB 12; Length 185;  
 Best Local Similarity 83.3%; Pred. No. 2.7;  
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cctctagagccagcagtcattgttgc 31  
 II III IIIIIIIIIIIIIIIII III  
 DB 80 CCGTCAGAACCCAGGATGTCATTGTACCC 51

RESULT 5  
 AAN70319/c  
 ID AAN70319 standard; DNA; 196 BP.  
 XX  
 AC AAN70319;  
 XX  
 DT 02-APR-1991 (first entry)  
 XX  
 DE Sequence encoding desulphatohirudin variant 1 (HIV).  
 XX  
 KW Anticoagulant; thrombin inhibitor; ds.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..196  
 FT /\*tag= a  
 FT  
 XX  
 PN EP225633-A.  
 XX  
 PD 16-JUN-1987.  
 XX  
 PF 09-DEC-1986; 86EP-0117098.  
 XX  
 PR 29-MAY-1986; 86GB-0013088.  
 PR 12-DEC-1985; 85GB-0030631.  
 XX  
 XX (CIBA ) CIBA GEIGY AG.  
 PA (PLAN-) PLANTORGAN WERK HEINRICH.  
 PA (CHRI-) PLANTORGANW CHRISTENSEN.  
 XX  
 PI Meyhack B, Marki W, Helm J;  
 XX  
 DR WPI; 1987-164868/24.  
 DR P-PSDB; AAP70225.  
 XX  
 XX New DNA constructs and hybrid vectors for transformation of yeast  
 PT etc., useful for prodn. and secretion of protein with hirudin  
 PT activity for use as thrombin inhibitors.  
 XX  
 XX Disclosure: pi8; 146pp; English.  
 XX  
 XX The preferred DNA construct of the invention contains the PHO5  
 CC promoter and a DNA segment consisting of the PHO5 signal sequence  
 CC upstream of and in reading frame with a DNA sequence coding for  
 CC mature desulphatohirudin. The segment is under the transcriptional  
 CC control of the PHO5 promoter and the 3' flanking sequence of the  
 CC PHO5 gene.  
 XX  
 XX Sequence 196 BP; 54 A; 49 C; 52 G; 41 T; 0 other;  
 SQ

Query Match 71.0%; Score 22; DB 8; Length 196;  
 Best Local Similarity 83.3%; Pred. No. 2.7;  
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cctctagagccagcagtcattgttgc 31  
 II III IIIIIIIIIIIIIIIII III

Db 101 CCGTCAGAACCCAGGATGTCATTGTTACCC 72

## RESULT 6

AAN60355/c  
ID AAN60355 standard; DNA; 217 BP.

XX AC AAN60355;

XX DT 20-JUN-1991 (first entry)

XX DE Desulphatohirudin (II).

XX KW Desulphatohirudin; antibodies; thrombin; ds.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX CDS 10..207

XX FT /\*tag= a

XX FT /product= hirudin

XX EP168342-A.

XX PD 15-JAN-1986.

XX PF 10-JUN-1985; 85EP-0810268.

XX PR 14-JUN-1984; 84CH-0288284.

XX PA (CIBA ) CIBA GEIGY AG.

XX LIersch M, Rink H, Marki W, Grutter MG, Meyhack B;

XX WPI; 1986-015589/03.

XX DR P-PSDB; AAP60395.

XX DNA sequences coding for hirudin and derivs. - and expression

XX PT vectors, transformed cells, monoclonal antibodies and hybridomas,

XX PT useful as thrombin inhibitor.

XX PS Disclosure; Page 10; 123pp; German.

XX CC The sequence comprises an EcoRI restriction enzyme site at the

XX CC 5'-end and a BamHI restriction enzyme site at the 3'-end.

XX CC The sequence may be introduced into a vector for the transformation

XX CC of hosts, e.g. E. coli. The hirudin can thus be prepared on a

XX CC large scale. The product and its derivs. are thrombin inhibitors,

XX CC useful in anticoagulant therapy, esp. when injected at doses

XX CC of 0.01-0.05 mg/kg.

XX SQ Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;

Query Match 71.0%; Score 22; DB 7; Length 217;

Best Local Similarity 83.3%; Pred. No. 2.7;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cctctagagccagatgcattgttgc 31

DB 113 CCGTCAGAACCCAGGATGTCATTGTTACCC 84

## RESULT 7

AAN70323/c  
ID AAN70323 standard; DNA; 217 BP.

XX AC AAN70323;

XX DT 02-APR-1991 (first entry)

XX DE Sequence of the desulphatohirudin gene.

XX DR P-PSDB; AAP60827.

KW Anticoagulant; thrombin inhibitor; ds.

XX EP225633-A.

XX PD 16-JUN-1987.

XX PF 09-DEC-1986; 86EP-0117098.

XX PR 29-MAY-1986; 86GB-0013088.

XX PR 12-DEC-1985; 85GB-0030631.

XX XX (CIBA ) CIBA GEIGY AG.

XX PA (PLAN-) PLANTORGAN WERK HEINRICH.

XX PA (CHRI-) PLANTORGANW CHRISTENSEN.

XX PI Meyhack B, Marki W, Heim J;

XX WPI; 1987-164868/24.

XX New DNA constructs and hybrid vectors for transformation of yeast

XX PT etc. - useful for prodn. and secretion of protein with hirudin

XX PT activity for use as thrombin inhibitors.

XX PS Example; p44; 146pp; English.

XX XX The preferred DNA construct of the invention contains the PHO5

XX CC promoter and a DNA segment consisting of the PHO5 signal sequence

XX CC upstream of and in reading frame with a DNA sequence coding for

XX CC mature desulphatohirudin. The segment is under the transcriptional

XX CC control of the PHO5 promoter and the 3' flanking sequence of the

XX CC PHO5 gene.

XX SQ Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;

Query Match

Best Local Similarity

Matches 25; Conservative

Mismatches 0; Indels

Gaps 0;

QY 2 cctctagagccagatgcattgttgc 31

DB 113 CCGTCAGAACCCAGGATGTCATTGTTACCC 84

## RESULT 8

AAN60746/c

ID AAN60746 standard; DNA; 224 BP.

XX AC AAN60746;

XX DT 01-JAN-1980 (first entry)

XX DE DNA encoding hirudin-like polypeptide.

XX KW Hirudin; thrombin-antagonist; anticoagulant; ds.

XX OS Synthetic.

XX XX DE3445517-A.

XX PD 19-JUN-1986.

XX PF 13-DEC-1984; 84DE-3445517.

XX PR 13-DEC-1984; 84DE-3445517.

XX PA (GENB-) GEN-BIO-TEC GES GEN.

XX PI Fortkamp E, Rieger M, Sommer R;

XX WPI; 1986-162802/26.

XX DR P-PSDB; AAP60827.







The sequence of the synthetaseirudin HW-1 genes was designed based on the published amino acid sequence (Dodd J., et al. *FEBS Letters* 165 180 (1984)). The sequence of streptokinase was obtained from PCR amplified chromosomal DNA from *S. equisimilis* ATCC 10009 or ATCC 9642. The primers used for the PCR were based on the published DNA sequence of *S. equisimilis* strain H46A (Walke, H., Roe, B., and Ferretti, J., *J. Gene* 34 357-362 (1985)). The two sequences were used to construct an expression vector in which the



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:31:51 ; Search time 174.99 Seconds<sup>3</sup>  
(without alignments)  
43.515 Million cell updates/sec

Title: US-10-053-641-6

Perfect score: 31

Sequence: 1 gcctctagagccaggatgcatgtgtgtgcc 31

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| C 1        | 26.2  | 84.5        | 235    | 1     | US-08-262-384A-7  |
| C 2        | 26.2  | 84.5        | 235    | 1     | US-08-262-384A-8  |
| C 3        | 26.2  | 84.5        | 726    | 1     | US-08-262-384A-9  |
| C 4        | 26.2  | 84.5        | 726    | 1     | US-08-262-384A-10 |
| C 5        | 25.4  | 81.9        | 59     | 1     | US-07-859-453E-9  |
| C 6        | 25.4  | 81.9        | 59     | 1     | US-07-859-453E-11 |
| C 7        | 25.4  | 80.6        | 129    | 1     | US-08-262-384A-3  |
| C 8        | 22.2  | 71.6        | 59     | 1     | US-07-859-453E-5  |
| C 9        | 22.2  | 71.0        | 217    | 6     | 5422249-13        |
| C 10       | 22.2  | 71.0        | 217    | 6     | 5422249-4         |
| C 11       | 22.2  | 71.0        | 279    | 1     | US-08-186-222-3   |
| C 12       | 21.8  | 70.3        | 64     | 6     | 5422249-11        |
| C 13       | 21.8  | 70.3        | 109    | 6     | 5422249-6         |
| C 14       | 20.8  | 67.1        | 40     | 1     | US-07-854-596B-54 |
| C 15       | 20.8  | 67.1        | 40     | 1     | US-07-854-596B-55 |
| C 16       | 20.8  | 67.1        | 201    | 1     | US-07-854-596B-1  |
| C 17       | 20.8  | 67.1        | 223    | 1     | US-07-854-596B-7  |
| C 18       | 20.8  | 67.1        | 420    | 1     | US-07-854-596B-8  |
| C 19       | 20.8  | 67.1        | 1458   | 1     | US-07-854-596B-42 |
| C 20       | 20.8  | 67.1        | 1467   | 1     | US-07-854-596B-46 |
| C 21       | 19.8  | 63.9        | 826    | 2     | US-08-560-098A-39 |
| C 22       | 19.8  | 63.9        | 826    | 2     | US-08-039-297B-7  |
| C 23       | 19.2  | 61.9        | 128    | 1     | US-08-262-384A-5  |
| C 24       | 19.2  | 61.9        | 212    | 6     | 5180668-10        |
| C 25       | 19.2  | 61.9        | 227    | 2     | US-07-982-064-8   |
| C 26       | 19.2  | 61.9        | 8491   | 2     | US-08-757-439-1   |
| C 27       | 19    | 61.3        | 59     | 1     | US-07-859-453E-7  |

|      |      |      |       |   |                   |                    |
|------|------|------|-------|---|-------------------|--------------------|
| 28   | 19   | 61.3 | 1817  | 4 | US-08-943-731-193 | Sequence 193, App  |
| 29   | 19   | 61.3 | 20084 | 4 | US-08-943-731-5   | Sequence 5, Appl   |
| C 30 | 18.8 | 60.6 | 198   | 1 | US-07-910-528-2   | Sequence 2, Appl   |
| C 31 | 18.8 | 60.6 | 198   | 1 | US-08-348-972-2   | Sequence 2, Appl   |
| C 32 | 18.8 | 60.6 | 227   | 6 | 5164304-3         | Patent No. 5164304 |
| C 33 | 18.8 | 60.6 | 228   | 6 | 5179196-3         | Patent No. 5179196 |
| C 34 | 18.8 | 60.6 | 2115  | 2 | US-08-767-026-3   | Sequence 3, Appl   |
| C 35 | 18.6 | 60.0 | 198   | 2 | US-08-861-459-2   | Sequence 2, Appl   |
| C 36 | 18.4 | 59.4 | 195   | 1 | US-08-044-506B-27 | Sequence 27, Appl  |
| C 37 | 18.4 | 59.4 | 1276  | 1 | US-08-417-460-7   | Sequence 7, Appl   |
| C 38 | 18.2 | 58.7 | 695   | 1 | US-08-592-126-80  | Sequence 80, Appl  |
| C 39 | 18.2 | 58.7 | 1269  | 1 | US-08-808-641-2   | Sequence 2, Appl   |
| C 40 | 18.2 | 58.7 | 1269  | 2 | US-09-064-839-2   | Sequence 2, Appl   |
| C 41 | 18.2 | 58.7 | 1289  | 3 | US-09-351-438-2   | Sequence 2, Appl   |
| C 42 | 17.8 | 57.4 | 195   | 1 | US-08-406-948A-7  | Sequence 7, Appl   |
| C 43 | 17.8 | 57.4 | 204   | 2 | US-08-492-343-1   | Sequence 1, Appl   |
| C 44 | 17.8 | 57.4 | 204   | 3 | US-09-080-865-2   | Sequence 2, Appl   |
| C 45 | 17.8 | 57.4 | 204   | 5 | PCT-US94-10048-1  | Sequence 1, Appl   |

## ALIGNMENTS

RESULT 1  
US-08-262-384A-7/c  
: Sequence 7, Application US/08262384A  
: Patent No. 5624822  
: GENERAL INFORMATION:  
: APPLICANT: Koerwer, Wolfgang  
: TITLE OF INVENTION: The Preparation of Hirudin  
: NUMBER OF SEQUENCES: 10  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Keil & Weinkauff  
: STREET: 1101 Connecticut Avenue  
: CITY: Washington  
: STATE: D.C.  
: COUNTRY: USA  
: ZIP: 20036  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
: OPERATING SYSTEM: IBM AT-compatible, 80486 processor  
: SOFTWARE: WordPerfect version 5.1  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/262,384A  
: FILING DATE: 20-JUN-1994  
: CLASSIFICATION: 530  
: CLASSIFICATION: C 12 N 15/62  
: CLASSIFICATION: C 12 N 15/31  
: CLASSIFICATION: C 07 K 7/10  
: CLASSIFICATION: C 12 P 21/02  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/EP90/02084  
: FILING DATE: 04-DEC-1990  
: APPLICATION NUMBER: US 07861820  
: FILING DATE: 18-JUN-1992  
: INFORMATION FOR SEQ ID NO: 7:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 235 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: linear  
: US-08-262-384A-7

Query Match 84.5%; Score 26.2; DB 1; Length 235;  
Best Local Similarity 90.3%; Pred. No. 0.004;  
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcctctagagccaggatgcatgtgtgtgcc 31  
||| |||||  
DB 135 GCGTCAGAGCCAGGATGCTTTGTGCC 105

```

RESULT 2
US-08-262-384A-8
; Sequence 8, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-262-384A-8

```

```

Query Match      84.5%; Score 26.2; DB 1; Length 235;
Best Local Similarity 90.3%; Pred. No. 0.004;
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 gctctagagccaggatgcattgtgtgcc 31
   ||| ||||| ||||| ||||| ||||| |||||
Db 105 GCCGTCAGAGCCAGGATGTCATTGTGTGCC 135

```

```

RESULT 3
US-08-262-384A-9/c
; Sequence 9, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/262,384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 726 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-262-384A-9

```

```

Query Match      84.5%; Score 26.2; DB 1; Length 726;
Best Local Similarity 90.3%; Pred. No. 0.005;
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 gctctagagccaggatgcattgtgtgcc 31
   ||| ||||| ||||| ||||| ||||| |||||
Db 627 GCCGTCAGAGCCAGGATGTCATTGTGTGCC 597

```

```

RESULT 4
US-08-262-384A-10
; Sequence 10, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 726 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-262-384A-10

```

```

Query Match      84.5%; Score 26.2; DB 1; Length 726;

```

Best Local Similarity 90.3%; Pred. No. 0.005;  
Matches 28; Conservative 0; Mismatches 3; Indels 3; Gaps 0;

QY 1 gctctagagccaggatgcattgtgtgcc 31  
Db 100 GCCGTCAGAGCCAGGATGATGTTGGCC 130

RESULT 5  
US-07-859-453E-9/c  
; Sequence 9, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:  
; APPLICANT: Kurfuerst, Manfred  
; APPLICANT: Ruebsamen, Klaus  
; APPLICANT: Schmied, Bernhard  
; APPLICANT: Koerwer, Wolfgang  
; APPLICANT: Schweden, Juergen  
; APPLICANT: Hoeffken, Hans Wolfgang  
; TITLE OF INVENTION: Hirudin/polyalkylene glycol  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS version 7.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/859,453E  
; FILING DATE: 29-MAY-1992  
; CLASSIFICATION: A 61 K 37/64  
; CLASSIFICATION: A 61 K 47/48  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/01998  
; FILING DATE: 22-NOV-1990  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 59 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown

US-07-859-453E-9

Query Match 81.9%; Score 25.4; DB 1; Length 59;  
Best Local Similarity 96.3%; Pred. No. 0.0071;  
Matches 26; Conservative 0; Mismatches 1; Indels 3; Gaps 0;

QY 1 gctctagagccaggatgcattgtgt 27  
Db 27 GCCTTAGAGCCAGGATGATGTT 1

RESULT 6  
US-07-859-453E-11/c  
; Sequence 11, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:  
; APPLICANT: Kurfuerst, Manfred  
; APPLICANT: Ruebsamen, Klaus  
; APPLICANT: Schmied, Bernhard  
; APPLICANT: Koerwer, Wolfgang  
; APPLICANT: Schweden, Juergen  
; APPLICANT: Hoeffken, Hans Wolfgang  
; TITLE OF INVENTION: Hirudin/polyalkylene glycol

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820

Query Match 81.9%; Score 25.4; DB 1; Length 59;  
Best Local Similarity 96.3%; Pred. No. 0.0071;  
Matches 26; Conservative 0; Mismatches 1; Indels 3; Gaps 0;

QY 1 gctctagagccaggatgcattgtgt 27  
Db 27 GCCTTAGAGCCAGGATGATGTT 1

RESULT 7  
US-08-262-384A-3/c  
; Sequence 3, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS version 7.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/859,453E  
; FILING DATE: 29-MAY-1992  
; CLASSIFICATION: 514  
; CLASSIFICATION: A 61 K 37/64  
; CLASSIFICATION: A 61 K 47/48  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/01998  
; FILING DATE: 22-NOV-1990  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 59 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown

US-07-859-453E-11

Query Match 81.9%; Score 25.4; DB 1; Length 59;  
Best Local Similarity 96.3%; Pred. No. 0.0071;  
Matches 26; Conservative 0; Mismatches 1; Indels 3; Gaps 0;

QY 1 gctctagagccaggatgcattgtgt 27  
Db 27 GCCTTAGAGCCAGGATGATGTT 1

RESULT 7  
US-08-262-384A-3/c  
; Sequence 3, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820

Query Match 81.9%; Score 25.4; DB 1; Length 59;  
Best Local Similarity 96.3%; Pred. No. 0.0071;  
Matches 26; Conservative 0; Mismatches 1; Indels 3; Gaps 0;

QY 1 gctctagagccaggatgcattgtgt 27  
Db 27 GCCTTAGAGCCAGGATGATGTT 1

RESULT 6  
US-07-859-453E-11/c  
; Sequence 11, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:  
; APPLICANT: Kurfuerst, Manfred  
; APPLICANT: Ruebsamen, Klaus  
; APPLICANT: Schmied, Bernhard  
; APPLICANT: Koerwer, Wolfgang  
; APPLICANT: Schweden, Juergen  
; APPLICANT: Hoeffken, Hans Wolfgang  
; TITLE OF INVENTION: Hirudin/polyalkylene glycol

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820

Query Match 81.9%; Score 25.4; DB 1; Length 59;  
Best Local Similarity 96.3%; Pred. No. 0.0071;  
Matches 26; Conservative 0; Mismatches 1; Indels 3; Gaps 0;

QY 1 gctctagagccaggatgcattgtgt 27  
Db 27 GCCTTAGAGCCAGGATGATGTT 1

RESULT 7  
US-08-262-384A-3/c  
; Sequence 3, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036



```

; APPLICANT: Schmitz, Albert
; TITLE OF INVENTION: Bacterial Vectors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/186,222
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,205
; FILING DATE: 19-MAR-1991
; APPLICATION NUMBER: GB 9006400.7
; FILING DATE: 22-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Villamizar, JoAnn
; REGISTRATION NUMBER: 30,598
; REFERENCE/DOCKET NUMBER: 4-17994/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914)785-7121
; TELEFAX: (914)347-5769
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (recombinant)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Hybrid gene: MSP signal (base pair 1 to 81)/
; INDIVIDUAL ISOLATE: desulfotetrudin (base pair 82 to 279)
; IMMEDIATE SOURCE:
; CLONE: pUCRS/pML310
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..276
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 82..276
; US-08-186-222-3

Query Match 71.0%; Score 22; DB 1; Length 279;
Best Local Similarity 83.3%; Pred. No. 0.33;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cctctagagccagatgcatgtgtgccc 31
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 CCGTCAGAACCCAGGATGTCATTTGTTACCC 153

RESULT 12
5422249-11
; Patent No. 5422249
; APPLICANT: LIERSCH, MANFRED; RINK, HANS; MARKI, WALTER; GRUTTER,
; MARKUS G.; MEYHACK, BERND
; TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF THROMBIN
; INHIBITORS
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/121,974
; FILING DATE: 15-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 700,9978
; FILING DATE: 10-MAY-1991
; APPLICATION NUMBER: 582,816
; FILING DATE: 13-SEP-1990
; APPLICATION NUMBER: 211,065
; FILING DATE: 20-JUN-1988
; APPLICATION NUMBER: 744,453
; FILING DATE: 13-JUN-1985
; SEQ ID NO: 6;
; LENGTH: 109
; 5422249-6

Query Match 70.3%; Score 21.8; DB 6; Length 109;
Best Local Similarity 92.0%; Pred. No. 0.34;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 agagccagagatgcatgtgtgccc 31
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 agaaccagagatgcatgtgttacc 26

RESULT 13
5422249-6/C
; Patent No. 5422249
; APPLICANT: LIERSCH, MANFRED; RINK, HANS; MARKI, WALTER; GRUTTER,
; MARKUS G.; MEYHACK, BERND
; TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF THROMBIN
; INHIBITORS
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/121,974
; FILING DATE: 15-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 700,9978
; FILING DATE: 10-MAY-1991
; APPLICATION NUMBER: 582,816
; FILING DATE: 13-SEP-1990
; APPLICATION NUMBER: 211,065
; FILING DATE: 20-JUN-1988
; APPLICATION NUMBER: 744,453
; FILING DATE: 13-JUN-1985
; SEQ ID NO: 6;
; LENGTH: 109
; 5422249-6

Query Match 70.3%; Score 21.8; DB 6; Length 109;
Best Local Similarity 92.0%; Pred. No. 0.34;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 agagccagagatgcatgtgtgccc 31
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 AGAACCCAGGATGTCATTTGTTACCC 84

RESULT 14
US-07-854-596B-54/C
; Sequence 54, Application US/07854596B
; Patent No. 5434073
; GENERAL INFORMATION:
; APPLICANT: Dawson, Keith M
; APPLICANT: Hunter, Michael G
; APPLICANT: Czaplowski, Lloyd G
; TITLE OF INVENTION: Proteins and nucleic acids
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. John J. McDonnell
; STREET: Ten South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/854,596B
FILING DATE: 03-JUN-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McDonnell, John J
REGISTRATION NUMBER: 26,949
REFERENCE/DOCKET NUMBER: 92,337
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 910-221-5317
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..40
OTHER INFORMATION: /note= "oligonucleotide BB2015"
US-07-854-596B-54

```

Query Match 67.1%; Score 20.8; DB 1; Length 40;  
Best Local Similarity 91.7%; Pred. No. 0.8;  
Matches 22; Conservative 0; Mismatches 2; Indels

RESULT 15  
US-07-854-596B-55  
; Sequence 55, Application US/07854596B  
; Patent No. 5434073  
; GENERAL INFORMATION:  
; APPLICANT: Dawson, Keith M  
; APPLICANT: Hunter, Michael G  
; APPLICANT: Czaplowski, Lloyd G  
; TITLE OF INVENTION: Proteins and nucleic acids  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dr. John J. McDonnell  
; STREET: Ten South Wacker Drive, Suite 3000  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/854,596B  
; FILING DATE: 03-JUN-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McDonnell, John J  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 92,337  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234

```

; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..40
; OTHER INFORMATION: /note= "oligonucleotide BB2016"
;
US-07-854-596B-55

Query Match 67.1%; Score 20.8; DB 1; Length 40;
Best Local Similarity 91.7%; Pred. No. 0.8;
Matches 22; Conservative 0; Mismatches 2; Indels

Qy 8 gagccaggatgcatttgtgcc 31
   ||| ||| ||| ||| ||| ||| |||
Db 16 GAACCCAGGATGCATTGTGACC 39

Search completed: July 15, 2002, 22:31:51
Job time: 18477 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:49:08 ; Search time 6165.88 Seconds  
(without alignments)  
67.858 Million cell updates/sec

Title: US-10-053-641-6

Perfect score: 31

Sequence: 1 gcctctagagccaggatgattgtgtccc 31

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hcc.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_hcc.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 24.2  | 78.1        | 459    | 10 | BI221953    |
| 2          | 24.2  | 78.1        | 555    | 10 | BI103096    |
| 3          | 22.6  | 72.9        | 197    | 9  | AW229748    |
| 4          | 22.6  | 72.9        | 586    | 10 | BI737971    |
| 5          | 21.2  | 68.4        | 463    | 12 | AQ664401    |
| c 6        | 21.2  | 67.7        | 658    | 12 | AG146871    |
| c 7        | 21.2  | 67.7        | 966    | 10 | BQ292991    |
| c 8        | 20.6  | 66.5        | 452    | 12 | AQ213776    |
| c 9        | 20.6  | 66.5        | 930    | 10 | BF683668    |
| 10         | 20.4  | 65.8        | 331    | 10 | BE708956    |
| 11         | 20.4  | 65.8        | 340    | 10 | BE696387    |
| c 12       | 20.4  | 65.8        | 357    | 10 | BF925749    |
| c 13       | 20.4  | 65.8        | 369    | 10 | BF195626    |
| 14         | 20.4  | 65.8        | 383    | 10 | BF745194    |
| 15         | 20.4  | 65.8        | 397    | 10 | BG876258    |
| c 16       | 20.4  | 65.8        | 401    | 10 | BF881067    |
| c 17       | 20.4  | 65.8        | 432    | 10 | W78212      |

|      |      |      |     |    |          |
|------|------|------|-----|----|----------|
| 18   | 20.4 | 65.8 | 445 | 10 | W30739   |
| 19   | 20.4 | 65.8 | 447 | 9  | AW408381 |
| 20   | 20.4 | 65.8 | 458 | 9  | AA812116 |
| 21   | 20.4 | 65.8 | 471 | 10 | W93251   |
| c 22 | 20.4 | 65.8 | 492 | 9  | AW406526 |
| 23   | 20.4 | 65.8 | 496 | 10 | BG385465 |
| 24   | 20.4 | 65.8 | 521 | 9  | AA662909 |
| 25   | 20.4 | 65.8 | 563 | 10 | BG116019 |
| 26   | 20.4 | 65.8 | 568 | 10 | BF219924 |
| 27   | 20.4 | 65.8 | 607 | 10 | BE795177 |
| c 28 | 20.4 | 65.8 | 650 | 9  | AA071197 |
| 29   | 20.4 | 65.8 | 675 | 9  | AW027138 |
| 30   | 20.4 | 65.8 | 711 | 10 | BI335232 |
| 31   | 20.4 | 65.8 | 715 | 12 | AG177881 |
| 32   | 20.4 | 65.8 | 726 | 10 | BG392336 |
| 33   | 20.4 | 65.8 | 757 | 10 | BE13262  |
| 34   | 20.4 | 65.8 | 761 | 10 | BE881549 |
| 35   | 20.4 | 65.8 | 789 | 10 | BI832624 |
| 36   | 20.4 | 65.8 | 848 | 9  | AU142879 |
| 37   | 20.4 | 65.8 | 855 | 10 | BM046549 |
| 38   | 20.4 | 65.8 | 891 | 10 | BI760539 |
| 39   | 20.4 | 65.8 | 895 | 10 | BE873926 |
| 40   | 20.4 | 65.8 | 898 | 10 | BG165109 |
| 41   | 20.4 | 65.8 | 901 | 10 | BE617436 |
| 42   | 20.4 | 65.8 | 901 | 10 | BE784393 |
| 43   | 20.4 | 65.8 | 924 | 10 | BG775094 |
| 44   | 20.4 | 65.8 | 941 | 10 | BG695997 |
| 45   | 20.4 | 65.8 | 964 | 10 | BE747898 |

## ALIGNMENTS

BI221953 602938496F1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:5101702 5', mRNA linear EST 11-JUL-2001  
mRNA sequence.  
BI221953  
BI221953.1 GI:14675397  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 459)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11244 row: p column: 23  
High quality sequence start: 7  
High quality sequence stop: 459.  
Location/Qualifiers  
1. 459  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5101702"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

## FEATURES

source

```

BASE COUNT      99 a 116 c 155 g 89 t
ORIGIN

Query Match      78.1%; Score 24.2; DB 10; Length 459;
Best Local Similarity 89.7%; Pred. No. 7.5;
Matches 26; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cctctagagccagagatgcatgttggcc 30
||||| | | | | | | | | | | | | | | | | |
Db 89 CCTCTACATCCAGGATGCAATTTGTGGCC 117

RESULT 2
BI103096
LOCUS      BI103096      555 bp      mRNA      linear      EST 26-JUN-2001
DEFINITION 602889325F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5044494
5', mRNA sequence.
ACCESSION  BI103096
VERSION     BI103096.1 GI:14553989
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 555)
/organism="Mus musculus"
/strain="CZECH II (feral)"
/db_xref="taxon:10090"
/clone_lib="IMAGE:5044494"
/clone_lib="NCI_CGAP_Kid14"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM1122 row: a column: 07
High quality sequence stop: 555.
Location/Qualifiers
1..555
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="IMAGE:5044494"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"

BASE COUNT      107 a 155 c 180 g 113 t
ORIGIN

Query Match      78.1%; Score 24.2; DB 10; Length 555;
Best Local Similarity 89.7%; Pred. No. 7.9;
Matches 26; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cctctagagccagagatgcatgttggcc 30
||||| | | | | | | | | | | | | | | | | |
Db 187 CCTCTACATCCAGGATGCAATTTGTGGCC 215

RESULT 3
AW229748
LOCUS      AW229748      197 bp      mRNA      linear      EST 10-DEC-1999
DEFINITION u041g07.y1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:2645148 5'
similar to TR:Q920G7 Q920G7 BING1.; mRNA sequence.
ACCESSION  AW229748
VERSION     AW229748.1 GI:6559044
KEYWORDS    EST.

```

```

SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 197)
/organism="Mus musculus"
/strain="CZECH II (feral)"
/db_xref="taxon:10090"
/clone_lib="IMAGE:2645148"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

MGI:1025600
Seq primer: -40RP from Gibco
High quality sequence stop: 181.
Location/Qualifiers
1..197
/organism="Mus musculus"
/strain="CZECH II (feral)"
/db_xref="taxon:10090"
/clone_lib="IMAGE:2645148"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium (MGC)

BASE COUNT      34 a 58 c 62 g 43 t
ORIGIN

Query Match      72.9%; Score 22.6; DB 9; Length 197;
Best Local Similarity 86.2%; Pred. No. 28;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cctctagagccagagatgcatgttggcc 30
||||| | | | | | | | | | | | | | | | | |
Db 52 CCTCTACATCCAGGATGCAATTTGTGGCC 80

RESULT 4
BI737971
LOCUS      BI737971      586 bp      mRNA      linear      EST 20-SEP-2001
DEFINITION 603357252F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5364517 5',
mRNA sequence.
ACCESSION  BI737971
VERSION     BI737971.1 GI:15714984
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 586)
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="IMAGE:5364517"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium (MGC)

BASE COUNT      107 a 155 c 180 g 113 t
ORIGIN

Query Match      78.1%; Score 24.2; DB 10; Length 555;
Best Local Similarity 89.7%; Pred. No. 7.9;
Matches 26; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cctctagagccagagatgcatgttggcc 30
||||| | | | | | | | | | | | | | | | | |
Db 187 CCTCTACATCCAGGATGCAATTTGTGGCC 215

RESULT 3
AW229748
LOCUS      AW229748      197 bp      mRNA      linear      EST 10-DEC-1999
DEFINITION u041g07.y1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:2645148 5'
similar to TR:Q920G7 Q920G7 BING1.; mRNA sequence.
ACCESSION  AW229748
VERSION     AW229748.1 GI:6559044
KEYWORDS    EST.

```

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLM11927 row: 9 column: 14

High quality sequence stop: 586.

#### FEATURES

source

```

1. 586
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5364517"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dr primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 115 a 167 c 180 g 124 t
ORIGIN

```

```

Query Match 72.9% Score 22.6; DB 10; Length 586;
Best Local Similarity 86.2% Pred. No. 37;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cctctagagccaggatgcattgtgccc 30
||||| |||||||||
Db 221 CCTCTACCTCCAGGATGCATTGTGGCC 249

```

#### RESULT 5

```

A0664401
LOCUS HS_5480_B1_B09_SP6E RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=1056 Col=17 Row=D, DNA sequence.
ACCESSION A0664401
VERSION A0664401.1 GI:5172169
KEYWORDS GSS:
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 463)
Mahairas G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1056 row: D column: 17
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 463.
Location/Qualifiers
1. 463
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1056 Col=17 Row=D"
/clone_lib="RPCI-11 Human Male BAC Library"

```

#### FEATURES

source

```

1. 463
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1056 Col=17 Row=D"
/clone_lib="RPCI-11 Human Male BAC Library"

```

/sex="male"

/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACE3.6 vector at EcoRI sites"

```

BASE COUNT 165 a 98 c 86 g 113 t 1 others
ORIGIN

```

```

Query Match 68.4% Score 21.2; DB 12; Length 463;
Best Local Similarity 85.2% Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 ctctagagccaggatgcattgttgc 29
||||| |||||||
Db 269 CACTTGAGCCCGCATATGTTGC 295

```

#### RESULT 6

```

AG146871/c
LOCUS Pan troglodytes DNA, clone: RP43-008J16.T7, genomic survey
DEFINITION sequence.
ACCESSION AG146871
VERSION AG146871.1 GI:16676549
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
Male BAC Library clone:RP43-008J16.T7.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (sites)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library RPCI-43
Unpublished
2 (bases 1 to 658)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: T7
LIBRARY
Vector : pBACE3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
Location/Qualifiers
1. 658
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-008J16.T7"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT 230 a 132 c 107 g 189 t
ORIGIN

```

#### COMMENT

Clones are derived from the chimpanzee BAC library RPCI-43 This BAC  
end was generated during the R&D process and may have higher chance  
of clone tracking errors.

#### PRIMERS

Sequencing: T7

Vector : pBACE3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI.

Location/Qualifiers

1. 658

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="RP43-008J16.T7"

/sex="male"

/cell\_type="lymphocytes"

/clone\_lib="RPCI-43 Chimpanzee Male BAC Library"

BASE COUNT 230 a 132 c 107 g 189 t

ORIGIN

#### Query Match

Best Local Similarity 67.7% Score 21; DB 12; Length 658;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cctctagagccaggatgcattgttgc 30

||||| ||||||| || ||||||| ||



into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCAGG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."

BASE COUNT  
 ORIGIN

244 a 331 c 250 g 105 t

Query Match 66.5%; Score 20.6; DB 10; Length 930;  
 Best Local Similarity 85.2%; Pred. No. 2.8e+02;  
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 cctctagagccagcagtcattgttg 28  
 ||||| ||||| ||||| ||||| |||||  
 Db 376 CCTCTAGTCCAGGCTCCATTGTTG 350

RESULT 10

BE708956  
 LOCUS BE708956 331 bp mRNA linear EST 12-SEP-2000  
 DEFINITION QV2-HT0577-160500-218-h08 HT0577 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE708956  
 VERSION BE708956.1 GI:10097221  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 331)

DIAS NETO, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922  
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl-st2-QV2-HT0577-160  
 500-218-h08st3-2000-05-16&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 56

High quality sequence stop: 197.

FEATURES  
 source

1. .331  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HT0577"  
 /dev\_stage="Adult"

/note="Organ: head\_neck; Vector: puc18; Site:1: SmaI;  
 Site:2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

BASE COUNT  
 ORIGIN

75 a 79 c 91 g 86 t

Query Match 65.8%; Score 20.4; DB 10; Length 331;  
 Best Local Similarity 80.0%; Pred. No. 2.6e+02;  
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gctctagagccagcagtcattgttgc 30  
 ||||| ||||| ||||| ||||| |||||  
 Db 176 GGCTCAGAACCCAGGAAGAAATTGTC 205

RESULT 11

BE596387  
 LOCUS BE596387 340 bp mRNA linear EST 11-SEP-2000  
 DEFINITION QV0-CT0383-230600-268-c04 CT0383 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE596387  
 VERSION BE596387.1 GI:10083547  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 340)

DIAS NETO, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922  
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl-st2-QV0-CT0383-230  
 600-268-c04st3-2000-06-23&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 9

High quality sequence stop: 338.

FEATURES  
 source

1. .340  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="CT0383"  
 /dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site:1: SmaI; Site:2:  
 SmaI; A mini-library was made by cloning products derived  
 from ORESTES PCR (U.S. Letters Patent application No. 196  
 ,716 - Ludwig Institute for Cancer Research) profiles  
 into the pUC 18 vector. Reverse transcription of tissue  
 mRNA and cDNA amplification were performed under low  
 stringency conditions."

BASE COUNT 93 a 81 c 89 g 77 t  
 ORIGIN

Query Match 65.8%; Score 20.4; DB 10; Length 340;  
 Best Local Similarity 80.0%; Pred. No. 2.6e+02;  
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gctctagagccagcagtcattgttgc 30  
 ||||| ||||| ||||| ||||| |||||  
 Db 172 GGCTCAGAACCCAGGAAGAAATTGTC 201



## Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-BT0859-011100-402-b09&t3=2000-11-01&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 31

High quality sequence stop: 100.

## FEATURES

source

1..383

Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BT0859"

/dev\_stage="Adult"

/note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2:

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles

into the pUC 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

96 a 104 c 99 g 84 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 65.8%; Score 20.4; DB 10; Length 383;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 gctcttagagccaggatgcttctgtgcc 30

||||| ||||| ||||| ||||| |||||

Db 72 GCTCTTAGAGCCAGGAGAAATTTGTTGCC 101

RESULT 15

LOCUS

BG876258 397 bp mRNA linear EST 30-MAY-2001

QV0-CT0383-150200-116-c06 CT0383 Homo sapiens cDNA, mRNA sequence.

ACCESSION

BG876258.1 GI:14253348

VERSION

EST.

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 397)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV0-CT0383-150

200-116-c06&t3=2000-02-15&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 5

High quality sequence stop: 397.

FEATURES

source

1..397

Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="CT0383"

/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2:

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles

into the pUC 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

109 a 100 c 103 g 85 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 65.8%; Score 20.4; DB 10; Length 397;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 gctcttagagccaggatgcttctgtgcc 30

||||| ||||| ||||| ||||| |||||

Db 167 GGCTCCAGACCCAGGAGAAATTTGTTGCC 196

Search completed: July 15, 2002, 21:49:15

Job time: 18906 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:29:29 ; Search time 2368.24 Seconds  
(without alignments)  
335.780 Million cell updates/sec

Title: US-10-053-641-7

Perfect score: 38

Sequence: 1 ggctctagagcgagaaatcaatcggttactggcga 38

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pi.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_hgt\_hum.\*

31: em\_hgt\_inv.\*

32: em\_hgt\_other.\*

33: em\_hgtc\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|----|-------------|
| -----      |       |             |           |    |             |

|    |      |      |       |    |          |
|----|------|------|-------|----|----------|
| 1  | 30   | 78.9 | 59    | 6  | I63542   |
| 2  | 30   | 78.9 | 59    | 6  | I63546   |
| 3  | 30   | 78.9 | 128   | 6  | I41185   |
| 4  | 30   | 78.9 | 226   | 12 | SYNHIR   |
| 5  | 30   | 78.9 | 235   | 6  | I41187   |
| 6  | 30   | 78.9 | 235   | 6  | I41188   |
| 7  | 30   | 78.9 | 726   | 6  | I41189   |
| 8  | 30   | 78.9 | 726   | 6  | I41190   |
| 9  | 28.4 | 74.7 | 59    | 6  | I63548   |
| 10 | 28   | 73.7 | 195   | 6  | I25751   |
| 11 | 28   | 73.7 | 215   | 6  | A36136   |
| 12 | 28   | 73.7 | 215   | 6  | E08760   |
| 13 | 26.8 | 70.5 | 563   | 6  | I25747   |
| 14 | 26.8 | 70.5 | 567   | 6  | A36139   |
| 15 | 25.8 | 67.9 | 106   | 6  | I41184   |
| 16 | 25.2 | 66.3 | 210   | 6  | E02086   |
| 17 | 25.2 | 66.3 | 279   | 6  | A01139   |
| 18 | 25.2 | 66.3 | 279   | 6  | I26534   |
| 19 | 25.2 | 66.3 | 526   | 6  | E02241   |
| 20 | 25.2 | 66.3 | 526   | 6  | E03810   |
| 21 | 24.4 | 64.2 | 201   | 6  | A34242   |
| 22 | 24.2 | 63.7 | 43    | 6  | A30667   |
| 23 | 24.2 | 63.7 | 43    | 6  | I38856   |
| 24 | 23.6 | 62.1 | 35    | 6  | A30668   |
| 25 | 23.6 | 62.1 | 35    | 6  | I38857   |
| 26 | 23.6 | 62.1 | 59    | 6  | I63544   |
| 27 | 23.6 | 62.1 | 67    | 6  | A34230   |
| 28 | 23.6 | 62.1 | 122   | 6  | A34236   |
| 29 | 23.6 | 62.1 | 122   | 6  | A34237   |
| 30 | 23.6 | 62.1 | 195   | 6  | A03693   |
| 31 | 23.6 | 62.1 | 195   | 6  | A03694   |
| 32 | 23.6 | 62.1 | 195   | 6  | A31866   |
| 33 | 23.6 | 62.1 | 195   | 6  | I62835   |
| 34 | 23.6 | 62.1 | 212   | 6  | A13382   |
| 35 | 23.6 | 62.1 | 212   | 6  | A13383   |
| 36 | 23.6 | 62.1 | 212   | 6  | A34619   |
| 37 | 23.6 | 62.1 | 212   | 6  | A34620   |
| 38 | 23.6 | 62.1 | 212   | 6  | E00711   |
| 39 | 23.6 | 62.1 | 217   | 6  | A03695   |
| 40 | 23.6 | 62.1 | 217   | 6  | A03696   |
| 41 | 23.6 | 62.1 | 217   | 6  | A34238   |
| 42 | 23.6 | 62.1 | 217   | 6  | A34239   |
| 43 | 23.6 | 62.1 | 217   | 6  | E00657   |
| 44 | 23.6 | 62.1 | 8491  | 6  | AR031529 |
| 45 | 23.4 | 61.6 | 73174 | 2  | AC036161 |

## ALIGNMENTS

|             |   |            |       |            |        |
|-------------|---|------------|-------|------------|--------|
| RESULT      | 1   |            |       |            |        |
| LOCUS       | I63542  | I63542     | 59 bp | DNA        | linear |
| DEFINITION  | Sequence 5 from patent US 5663141.  |            |       |            |        |
| ACCESSION   | I63542  |            |       |            |        |
| VERSION     | I63542.1  | GI:2481115 |       |            |        |
| KEYWORDS    |   |            |       |            |        |
| SOURCE      | Unknown.  |            |       |            |        |
| ORGANISM    | Unknown.  |            |       |            |        |
| REFERENCE   | 1 (bases 1 to 59)   |            |       |            |        |
| AUTHORS     | Kurfuerst, M., Ruebsamen, K., Schmied, B., Koerwer, W., Schweden, J. and Hoeftken, H. Wolfgang. |            |       |            |        |
| TITLE       | Hirudin/polysaccharide glycol conjugates and hirudin muteins                                    |            |       |            |        |
| JOURNAL     | Patent: US 5663141-A 5 02-SEP-1997;   |            |       |            |        |
| FEATURES    | Location/Qualifiers   |            |       |            |        |
| source      | 1..59   |            |       |            |        |
| BASE COUNT  | 19 a  | 14 c       | 16 g  | 10 t       |        |
| ORIGIN      |   |            |       |            |        |
| Query Match | 78.9%   | Score 30;  | DB 6; | Length 59; |        |

Best Local Similarity 86.8%; Pred. No. 0.052; Mismatches 5; Indels 0; Gaps 0;  
Matches 33; Conservative 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38  
||||| ||||||| || ||||||| |||||  
Db 16 GGCTCTGACGGCGAAAAAACCCAGTGGTACTGGCGA 53

RESULT 2  
LOCUS I63546 59 bp DNA PAT 07:OCT-1997  
DEFINITION Sequence 9 from patent US 5663141.  
ACCESSION I63546  
VERSION I63546.1 GI:2481119  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS Kurfuerst, M., Ruebsamen, K., Schmied, B., Koerwer, W., Schweden, J. and  
Hoeffken, H. Wolfgang.  
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins  
JOURNAL Patent: US 5663141-A 9 02-SEP-1997;  
FEATURES Location/Qualifiers  
source 1..59  
BASE COUNT 20 a 14 c 15 g 10 t  
ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 59;  
Best Local Similarity 86.8%; Pred. No. 0.052;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38  
||||| ||||||| || ||||||| |||||  
Db 16 GGCTCTAAAGCGCGAAAAAACCCAGTGGTACTGGCGA 53

RESULT 3  
LOCUS I41185 128 bp DNA PAT 13-MAY-1997  
DEFINITION Sequence 5 from patent US 5624822.  
ACCESSION I41185  
VERSION I41185.1 GI:2081775  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 5 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..128  
BASE COUNT 20 a 33 c 34 g 41 t  
ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 128;  
Best Local Similarity 86.8%; Pred. No. 0.054;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38  
||||| ||||||| || ||||||| |||||  
Db 116 GGCTCTGACGGCGAAAAAACCCAGTGGTACTGGCGA 79

RESULT 4  
LOCUS SYNHIR 226 bp DNA SYN 27-APR-1993  
DEFINITION Synthetic hirudin gene, complete cds.

ACCESSION M26762  
VERSION M26762.1 GI:208478  
KEYWORDS hirudin; proteinase inhibitor; thrombin inhibitor.  
SOURCE Synthetic DNA.  
ORGANISM Synthetic construct  
artificial sequence.  
REFERENCE 1 (bases 1 to 226)  
AUTHORS Bergmann, C., Dodd, J., Koehler, S., Fink, E. and Gassen, H. G.  
TITLE Chemical synthesis and expression of a gene coding for hirudin, the  
thrombin-specific inhibitor from the leech Hirudo medicinalis  
JOURNAL Biol. Chem. Hoppe-Seyler 367, 731-740 (1986)  
MEDLINE 87026239  
FEATURES Location/Qualifiers  
source 1..226  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
CDS 22..222  
/note="hirudin"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AAA72772.1"  
/db\_xref="GI:208478"  
/translation="MVFYDCTESGONLCLEGSNVGOGNKCILGSDGKNOQVGTGE  
GTFKPSHNDGDFEIEPEYLQ"  
BASE COUNT 63 a 56 c 61 g 46 t  
ORIGIN

Query Match 78.9%; Score 30; DB 12; Length 226;  
Best Local Similarity 86.8%; Pred. No. 0.056;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38  
||||| ||||||| || ||||||| |||||  
Db 115 GGCTCTGACGGCGAAAAAACCCAGTGGTACTGGCGA 152

RESULT 5  
LOCUS I41187 235 bp DNA PAT 13-MAY-1997  
DEFINITION Sequence 7 from patent US 5624822.  
ACCESSION I41187  
VERSION I41187.1 GI:2081777  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 7 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..235  
/organism="unknown"  
BASE COUNT 72 a 63 c 57 g 43 t  
ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 235;  
Best Local Similarity 86.8%; Pred. No. 0.056;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38  
||||| ||||||| || ||||||| |||||  
Db 124 GGCTCTGACGGCGAAAAAACCCAGTGGTACTGGCGA 161

RESULT 6  
LOCUS I41188/c 235 bp DNA PAT 13-MAY-1997  
DEFINITION Sequence 8 from patent US 5624822.  
ACCESSION I41188  
VERSION I41188.1 GI:2081778

KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 235)

AUTHORS Koerwer, W.

TITLE Hirudin fusion proteins and preparation of hirudin

JOURNAL Patent: US 5624822-A 8 29-APR-1997;

FEATURES Location/Qualifiers

1..235

source /organism="unknown"

BASE COUNT 42 a 58 c 64 g 71 t

ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 235;  
Best Local Similarity 86.8%; Pred. No. 0.056;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgagaaaaaatcaatcggttactggcga 38

Db 116 GGCTCTGACGGCGAAGAAAAACCAAGTCGTTACTGGCGA 79

RESULT 7

I41189

LOCUS I41189

DEFINITION Sequence 9 from patent US 5624822.

ACCESSION I41189

VERSION I41189.1 GI:2081779

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 726)

AUTHORS Koerwer, W.

TITLE Hirudin fusion proteins and preparation of hirudin

JOURNAL Patent: US 5624822-A 9 29-APR-1997;

FEATURES Location/Qualifiers

1..726

source /organism="unknown"

BASE COUNT 281 a 171 c 126 g 148 t

ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 726;  
Best Local Similarity 86.8%; Pred. No. 0.059;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgagaaaaaatcaatcggttactggcga 38

Db 616 GGCTCTGACGGCGAAGAAAAACCAAGTCGTTACTGGCGA 653

RESULT 8

I41190/c

LOCUS I41190

DEFINITION Sequence 10 from patent US 5624822.

ACCESSION I41190

VERSION I41190.1 GI:2081780

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 726)

AUTHORS Koerwer, W.

TITLE Hirudin fusion proteins and preparation of hirudin

JOURNAL Patent: US 5624822-A 10 29-APR-1997;

FEATURES Location/Qualifiers

1..726

source /organism="unknown"

BASE COUNT 149 a 126 c 171 g 280 t

ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 726;  
Best Local Similarity 86.8%; Pred. No. 0.059;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgagaaaaaatcaatcggttactggcga 38

Db 111 GGCTCTGACGGCGAAGAAAAACCAAGTCGTTACTGGCGA 74

RESULT 9

I63548

LOCUS I63548

DEFINITION Sequence 11 from patent US 5663141.

ACCESSION I63548

VERSION I63548.1 GI:2481121

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 59)

AUTHORS Hoeffken, H. Wolfgang.

TITLE Hirudin/polyalkylene glycol conjugates and hirudin mutants

JOURNAL Patent: US 5663141-A 11 02-SEP-1997;

FEATURES Location/Qualifiers

1..59

source /organism="unknown"

BASE COUNT 18 a 14 c 16 g 11 t

ORIGIN

Query Match 74.7%; Score 28.4; DB 6; Length 59;  
Best Local Similarity 84.2%; Pred. No. 0.24;  
Matches 32; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ggctctagagcgagaaaaaatcaatcggttactggcga 38

Db 16 GGCTCTAAGCGGAACGTAACCAAGTCGTTACTGGCGA 53

RESULT 10

I25751

LOCUS I25751

DEFINITION Sequence 27 from patent US 5552299.

ACCESSION I25751

VERSION I25751.1 GI:1605621

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 195)

AUTHORS Ott, Ian., Klupp, T., Moln ar, Ian., Pathy, Aas., Barta, Ian., Bark o n

ee T oth, Z., Ambrus, Gabor., Sal at, Janos., Tegdes, Ao.,

Moravcsik, I., Egy ud, C., Albrecht, Karly., K oncz Ol, Kalman.,

Vince, A., Barab as, E., M at e, Gorgy., Kiss, Gorgy. B., Kiss, Peter.,

P olya, Kalman., Erdei, Janos., Guly as, E. and Zilahi, E.

Plasmids and process for producing recombinant desulphatohirudin

HV-1 peptides

JOURNAL Patent: US 5552299-A 27 03-SEP-1996;

FEATURES Location/Qualifiers

1..195

source /organism="unknown"

BASE COUNT 58 a 40 c 45 g 51 t

ORIGIN

Query Match 73.7%; Score 28; DB 6; Length 195;  
Best Local Similarity 81.6%; Pred. No. 0.37;  
Matches 31; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 ggctctagagcgagaaaaaatcaatcggttactggcga 38

DB 91 GGTCTACGGTGAAAAAATCAATGTGTCTACTGCGCA 128  
 || ||||: || |||||||||||||||| || |||||||||  
 RESULT 11  
 A36136 215 bp DNA linear PAT 04-MAR-1997  
 DEFINITION Sequence 1 from Patent EP0576792.  
 ACCESSION A36136  
 VERSION A36136.1 GI:2293652  
 KEYWORDS  
 SOURCE Helobdella triserialis.  
 ORGANISM Helobdella triserialis  
 Helobdella triserialis  
 Eukaryota; Metazoa; Annelida; Clitellata; Hirudinea; Hirudinea;  
 Rhynchobdellida; Glossiphoniidae; Helobdella.  
 REFERENCE 1 (bases 1 to 215)  
 AUTHORS Ott,I.D., Klupp,T.D., Molnar,I., Patthy,A.D., Barta,I.,  
 Berk-T.T., Ambrus,G.D., Salati,J., Tegdes,A., Moravcsik,I.I.,  
 Egyued,C., Albrecht,K., Koencsoel,K.D., Vincze,A., Barabas,E.D.,  
 Mate,G.I., Botond,K.G., Kiss,P., Poelya,K.D., Erdei,J.D.,  
 Erdei,J.S. and Zilahi,E.  
 TITLE Plasmids and process for producing recombinant desulphatohirudin  
 JOURNAL HV-1 peptide  
 BIOLOGICAL GYOGYSZERGYAR (HU)  
 FEATURES  
 SOURCE 1..215  
 Location/Qualifiers  
 /organism="Helobdella triserialis"  
 /db\_xref="taxon:6413"  
 BASE COUNT 63 a 44 c 51 g 56 t 1 others  
 ORIGIN  
 Query Match 73.7%; Score 28; DB 6; Length 215;  
 Best Local Similarity 81.6%; Pred. No. 0.37;  
 Matches 31; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 ggctctagaggcgaaaaaatcaatgcgttactgcgca 38  
 || ||||: || |||||||||||||||| || |||||||||  
 DB 100 GGTCTACGGTGAAAAAATCAATGTGTCTACTGCGCA 137  
 || ||||: || |||||||||||||||| || |||||||||  
 RESULT 12  
 E08760 215 bp DNA linear PAT 29-SEP-1997  
 LOCUS  
 DEFINITION DNA with Escherichia coli-Saccharomyces codon which encodes  
 desulphatohirudin,HV-1 peptide.  
 ACCESSION E08760  
 VERSION E08760.1 GI:2176872  
 KEYWORDS JP 1995046995-A/1.  
 SOURCE unidentified.  
 ORGANISM unclassified.  
 REFERENCE 1 (bases 1 to 215)  
 AUTHORS Ishiyutobuan,O., Teiboru,K., Ishiyutobuan,M., Andoraashiyu,P.,  
 Ishiyutobuan,B., Jiyujiya,B.N.T., Gaaboru,A., Yaanoshiyu,S.,  
 Anikoo,T., Imure,M., Tsuetsuira,E., Kaarori,A., Kaarumaan,K.,  
 Atsuteira,B., Eebua,B., Jiyoeurujii,M., Jiyoeurujii,B.K., Peeteru,K.,  
 Kaarumaan,P., Yaanoshiyu,E., Eebua,G. and Erika,Z.  
 TITLE PREPARATION OF PLASIMID AND RECOMBINANT DESULFATOHIRUDINE HV-1  
 JOURNAL Patent: JP 1995046995-A 1 21-FEB-1995;  
 BIOLOGICAL GYOGYSZERGYAR  
 COMMENT OS None  
 OC Artificial sequences.  
 PN JP 1995046995-A/1  
 PD 21-FEB-1995  
 PF 09-APR-1993 JP 1993105996  
 PI ISHIYUTOBUAN OTSUTO, TEIBORU KURUTSUPU,  
 PI ISHIYUTOBUAN MORUNARU,  
 PI ANDORAASHIYU PATSUTSUII, ISHIYUTOBUAN BARUTA, PI JIYUJIYA  
 BARUKO NEE TOTSU, GAABORU AMUBURUSHIYU, PI YAANOSHIYU  
 SARAATOU,  
 PI ANIKOO TEGUDESIIYU, IMURE MORABUCHITSUKU, TSUETSUIRA EJIYUDO,  
 PI

DB 91 GGTCTACGGTGAAAAAATCAATGTGTCTACTGCGCA 128  
 || ||||: || |||||||||||||||| || |||||||||  
 RESULT 11  
 A36136 215 bp DNA linear PAT 04-MAR-1997  
 DEFINITION Sequence 1 from Patent EP0576792.  
 ACCESSION A36136  
 VERSION A36136.1 GI:2293652  
 KEYWORDS  
 SOURCE Helobdella triserialis.  
 ORGANISM Helobdella triserialis  
 Helobdella triserialis  
 Eukaryota; Metazoa; Annelida; Clitellata; Hirudinea; Hirudinea;  
 Rhynchobdellida; Glossiphoniidae; Helobdella.  
 1 (bases 1 to 215)  
 REFERENCE (Ott,I.D., Klupp,T.D., Molnar,I., Patthy,A.D., Barta,I.,  
 AUTHORS Berk-T.T., Ambrus,G.D., Salati,J., Tegdes,A., Moravcsik,I.I.,  
 Egyued,C., Albrecht,K., Koencsoel,K.D., Vincze,A., Barabas,E.D.,  
 Mate,G.I., Botond,K.G., Kiss,P., Poelya,K.D., Erdei,J.D.,  
 Erdei,J.S. and Zilahi,E.  
 Plasmids and process for producing recombinant desulphatohirudin  
 HV-1 peptide  
 JOURNAL Patent: EP 0576792-A 1 05-JAN-1994;  
 BIOGAL GYOGYSZERGYAR (HU)  
 FEATURES Location/Qualifiers  
 source 1..215  
 /organism="Helobdella triserialis"  
 /db\_xref="taxon:6413"  
 BASE COUNT 63 a 44 c 51 g 56 t 1 others  
 ORIGIN  
 Query Match 73.7%; Score 28; DB 6; Length 215;  
 Best Local Similarity 81.6%; Pred. No. 0.37;  
 Matches 31; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 ggctctagaggcgaaaaaatcaatgcgttactgcgca 38  
 || ||||: || |||||||||||||||| || |||||||||  
 Db 100 GGTCTACGGTGAAAAAATCAATGTGTCTACTGCGCA 137  
 RESULT 12  
 E08760 215 bp DNA linear PAT 29-SEP-1997  
 LOCUS  
 DEFINITION DNA with Escherichia coli-Saccharomyces codon which encodes  
 desulphatohirudin,HV-1 peptide.  
 ACCESSION E08760  
 VERSION E08760.1 GI:2176872  
 KEYWORDS JP 1995046995-A/1.  
 SOURCE unidentified.  
 ORGANISM unclassified.  
 REFERENCE 1 (bases 1 to 215)  
 AUTHORS Ishiyutobuan,O., Teiboru,K., Ishiyutobuan,M., Andoraashiyu,P.,  
 Ishiyutobuan,B., Jiyujiya,B.N.T., Gaaboru,A., Yaanoshiyu,S.,  
 Anikoo,T., Imure,M., Tsuetsuira,E., Kaarori,A., Kaarumaan,K.,  
 Atsuteira,B., Eebua,B., Jiyoeurujii,M., Jiyoeurujii,B.K., Peeteru,K.,  
 Kaarumaan,P., Yaanoshiyu,E., Eebua,G. and Erika,Z.  
 PREPARATION OF PLASIMID AND RECOMBINANT DESULFATOHIRUDINE HV-1  
 JOURNAL Patent: JP 1995046995-A 1 21-FEB-1995;  
 BIOGAL GYOGYSZERGYAR  
 OS None  
 COMMENT OC Artificial sequences.  
 PN JP 1995046995-A/1  
 PD 21-FEB-1995  
 PF 09-APR-1993 JP 1993105996  
 PI ISHIYUTOBUAN OTSUTO, TEIBORU KURUTSUPU,  
 PI ISHIYUTOBUAN MORUNARU,  
 PI ANDORAASHIYU PATSUTSUII, ISHIYUTOBUAN BARUTA, PI JIYUJIYA  
 BARUKOO NEEB TOOTSU, GAABORU AMUBURUSHIYU, PI YAANOSHIYU  
 SARAATOU,  
 PI ANIKOO TEGUDESHIIYU, IMURE MORABUCHITSUKU, TSUETSUIRA EJIYUDO,

Search completed: July 15, 2002, 22:29:30  
Job time: 19016 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:55 ; Search time 758.37 Seconds  
(without alignments)  
86.030 Million cell updates/sec

Title: US-10-053-641-7  
Perfect score: 38  
Sequence: 1 ggctctagaggcgaaataatcaatgcgttactgscga 38

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

| Database :         |   |  |  |
|--------------------|---|--|--|
| N_Geneseq_032802.* |   |  |  |
| 1:                 | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*  |  |  |
| 2:                 | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*  |  |  |
| 3:                 | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*  |  |  |
| 4:                 | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*  |  |  |
| 5:                 | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*  |  |  |
| 6:                 | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*  |  |  |
| 7:                 | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*  |  |  |
| 8:                 | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*  |  |  |
| 9:                 | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*  |  |  |
| 10:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*  |  |  |
| 11:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*  |  |  |
| 12:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*  |  |  |
| 13:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*  |  |  |
| 14:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*  |  |  |
| 15:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*  |  |  |
| 16:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*  |  |  |
| 17:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*  |  |  |
| 18:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*  |  |  |
| 19:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*  |  |  |
| 20:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*  |  |  |
| 21:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*  |  |  |
| 22:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.* |  |  |
| 23:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.* |  |  |
| 24:                | /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*  |  |  |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 30    | 78.9        | 726    | 12 | AAQ12379    |
| 2          | 28.4  | 74.7        | 468    | 16 | AAQ93244    |
| 3          | 28    | 71.9        | 219    | 15 | AAQ54995    |
| 4          | 26.8  | 70.5        | 567    | 15 | AAQ54997    |
| 5          | 25.2  | 66.3        | 195    | 13 | AAQ27601    |
| 6          | 25.2  | 66.3        | 208    | 10 | AAQ91851    |
| 7          | 25.2  | 66.3        | 276    | 13 | AAQ27600    |
| 8          | 25.2  | 66.3        | 279    | 12 | AAQ13831    |
| 9          | 25.2  | 66.3        | 525    | 11 | AAQ06873    |

|    |      |      |      |    |          |                    |
|----|------|------|------|----|----------|--------------------|
| 10 | 25.2 | 66.3 | 526  | 11 | AAQ03237 | Sequence of DNA in |
| 11 | 24.4 | 64.2 | 231  | 7  | AAQ60354 | Desulphatohirudin  |
| 12 | 24.2 | 63.7 | 43   | 14 | AAQ43368 | Hirudin oligonucle |
| 13 | 23.6 | 62.1 | 35   | 14 | AAQ43369 | Hirudin oligonucle |
| 14 | 23.6 | 62.1 | 195  | 16 | AAQ81528 | Leech hirudin mute |
| 15 | 23.6 | 62.1 | 196  | 8  | AAQ70319 | Sequence encoding  |
| 16 | 23.6 | 62.1 | 217  | 7  | AAQ60355 | Desulphatohirudin  |
| 17 | 23.6 | 62.1 | 217  | 8  | AAQ70323 | Sequence of the de |
| 18 | 23.6 | 62.1 | 227  | 22 | AAQ61507 | S. marcescens hiru |
| 19 | 22.2 | 58.4 | 2445 | 19 | AAQ30596 | H. pylori outer me |
| 20 | 22.2 | 58.4 | 2445 | 20 | AAQ75781 | H. pylori outer me |
| 21 | 22.2 | 58.4 | 2448 | 19 | AAQ30595 | H. pylori outer me |
| 22 | 22.2 | 58.4 | 2448 | 20 | AAQ75780 | Partial Hirudin HV |
| 23 | 22   | 57.9 | 183  | 12 | AAQ14927 | Synthetic hirudin  |
| 24 | 22   | 57.9 | 185  | 12 | AAQ14926 | Sequence encoding  |
| 25 | 22   | 57.9 | 198  | 13 | AAQ24986 | Hirudin variant.   |
| 26 | 22   | 57.9 | 198  | 15 | AAQ63876 | HV-1 gene. Synthe  |
| 27 | 22   | 57.9 | 210  | 12 | AAQ10452 | DNA encoding hirud |
| 28 | 22   | 57.9 | 224  | 7  | AAQ60746 | Sequence encoding  |
| 29 | 22   | 57.9 | 229  | 9  | AAQ81294 | Sequence encoding  |
| 30 | 22   | 57.9 | 238  | 9  | AAQ81295 | Synthetic HVI gene |
| 31 | 22   | 57.9 | 257  | 10 | AAQ91836 | Synthetic hirudin  |
| 32 | 22   | 57.9 | 257  | 10 | AAQ91867 | CUP1 promoter, PHO |
| 33 | 22   | 57.9 | 358  | 6  | AAQ50397 | Yeast CUP1 promote |
| 34 | 22   | 57.9 | 1082 | 15 | AAQ64146 | pJDB207/GAPFL-YHIR |
| 35 | 22   | 57.9 | 1082 | 15 | AAQ45312 | Salmonella typhi D |
| 36 | 22   | 57.9 | 1130 | 16 | AAQ81527 | Human CDNA encodin |
| 37 | 21   | 55.3 | 1773 | 23 | AAQ56019 | DNA encoding novel |
| 38 | 20.8 | 54.7 | 745  | 22 | AAQ33871 | Synthetic hirudin  |
| 39 | 20.6 | 54.2 | 3884 | 23 | AAQ90132 | Hirudin HV-1. Syn  |
| 40 | 20.4 | 53.7 | 201  | 12 | AAQ12153 | Desulphatohirudin  |
| 41 | 20.4 | 53.7 | 201  | 13 | AAQ54996 | Factor Xa-cleavabl |
| 42 | 20.4 | 53.7 | 304  | 15 | AAQ25184 | Factor Xa-cleavabl |
| 43 | 20.4 | 53.7 | 420  | 12 | AAQ12155 | Factor Xa-cleavabl |
| 44 | 20.4 | 53.7 | 1458 | 12 | AAQ12162 | Factor Xa-cleavabl |
| 45 | 20.4 | 53.7 | 1467 | 12 | AAQ12490 |                    |

ALIGNMENTS

|          |  |
|----------|--|
| RESULT 1 |  |
| AAQ12379 |  |
| ID       | AAQ12379 standard; DNA; 726 BP.        |
| XX       |  |
| XX       |  |
| AC       | AAQ12379;                              |
| XX       |  |
| DT       | 17-SEP-1991 (first entry)              |
| XX       |  |
| DE       | Hirudin peptide/Protein A fusion gene. |
| XX       |  |
| XX       | anticoagulant; fusion protein; ds.     |
| KW       |  |
| XX       |  |
| PH       | Key                                    |
| FT       | mat_peptide                            |
| FT       | Location/Qualifiers                    |
| FT       | 1.489                                  |
| FT       | /*tag= a                               |
| FT       | /product= Protein A                    |
| FT       | 529..720                               |
| FT       | /*tag= b                               |
| FT       | /product= Hirudin                      |
| FT       | 490..528                               |
| FT       | /*tag= c                               |
| FT       | /product= linker oligopeptide Y        |
| XX       |  |
| PN       | DE3942580-A.                           |
| XX       |  |
| PD       | 27-JUN-1991.                           |
| XX       |  |
| PF       | 22-DEC-1989;                           |
| XX       | 89DE-3942580.                          |
| PR       | 22-DEC-1989;                           |
| XX       | 89DE-3942580.                          |
| PA       | (BADI ) BASF AG.                       |

```

XX PI Korwer W;
XX XX
XX DR WPI; 1991-194236/27.
XX DR P-PSDB; AAR12751.
XX XX
XX XX Hirudin peptide prodn. by cleaving new fusion peptide - of
XX PT hirudin and protein A, expressed in high yield and stable,
XX PT soluble form by transformed E.coli
XX XX
XX PS Example 1; Page 6-7; 9pp; German.
XX CC This sequence is an example of a fusion construct for expression of
XX CC the fusion peptide of the invention. The Met residue in the linker
XX CC oligopeptide allows cleavage by CNBr to release two fragments which
XX CC can be easily separated by IGC affinity chromatography. (The protein
XX CC A component binds to Igc sepharose). Increased yields of hirudin are
XX CC obtained using this fusion construct.
XX XX
XX SQ Sequence 726 BP; 280 A; 171 C; 126 G; 149 T; 0 other;

Query Match 78.9%; Score 30; DB 12; Length 726;
Best Local Similarity 86.8%; Pred. No. 0.0037;
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagaggcgagaaataaatcgcttactggcgga 38
      ||||| ||||| ||||| ||||| ||||| |||||
Db 616 ggctctagaggcgagaaataaatcgcttactggcgga 653

RESULT 2
AAQ93244
ID AAQ93244 standard; DNA; 468 BP.
XX AC AAQ93244;
XX XX
XX DT 01-NOV-1995 (first entry)
XX DE Fusion construct of glucoamylase-hirudin DNA.
XX KW expression cassette; recombinant protein; production; Hansenula;
XX KW yeast; leader sequence; adaptor; alpha helix; glucoamylase;
XX KW secretion; processing; thrombin inhibitor; hirudin; ds.
XX OS Synthetic.
XX XX
XX FH Key Location/Qualifiers
XX FT primer_bind complement (1..33)
XX FT /*tag= a
XX FT /note= "primer AAQ85840 binding site"
XX FT CDS 13..444
XX FT /*tag= b
XX FT /product= glucoamylase-hirudin fusion protein
XX FT 199..223
XX FT /*tag= c
XX FT /note= "primer AAQ85841 binding site"
XX FT primer_bind complement (220..267)
XX FT /*tag= d
XX FT /note= "primer AAQ85842 binding site"
XX FT misc_feature 13..228
XX FT /*tag= e
XX FT /note= "encodes amino acids 1-72 of glucoamylase"
XX FT misc_feature 241..246
XX FT /*tag= f
XX FT /note= "encodes processor signal"
XX FT misc_feature 247..444
XX FT /*tag= g
XX FT /note= "encodes hirudin-h120"
XX FT primer_bind 0..468
XX FT /*tag= h
XX FT /note= "primer AAQ85843 binding site"
XX XX

```

```

PN DE4329969-A.
XX PD
XX PD 09-MAR-1995.
XX PF
XX PF 04-SEP-1993; 93DE-4329969.
XX XX
XX XX 04-SEP-1993; 93DE-4329969.
XX XX
XX XX (BADI ) BASF AG.
XX PA (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
XX XX
XX XX Bollschweiler C, Janowicz ZA, Piontek M, Schweden J;
XX PI Strasserawm, Weydemann U;
XX XX
XX DR WPI; 1995-107810/15.
XX DR P-PSDB; AAR76951.
XX XX
XX PT Recombinant protein prodn. in Hansenula yeast - transformed with
XX PT expression cassette contg. leader, adaptor, processing signal and
XX PT gene, provides efficient secretion and correct processing
XX PS
XX PS Example 1; Fig 1; 10pp; German.
XX XX
XX CC AAQ93244 is a DNA fusion product resulting from ligation of an
XX CC EcoRI-PvuI fragment (see AAQ85840-41) and a PvuI-SalI fragment, encoding
XX CC a leader sequence, an adaptor (see AAR71472, amino acids 23-72 of GAM
XX CC (glucamylase from Schwanniomycetes occidentalis, plus -His-Pro-Leu-Gln at
XX CC the C-terminal end), a processor peptide (Lys-Arg) and a structural gene,
XX CC encoding hirudin. AAQ93244 is an example of an expression vector insert
XX CC of the general formula: L-A-P-Gene (L = leader sequence; A = adaptor;
XX CC P = processor; Gene = structural gene). The cassettes ensure efficient
XX CC secretion and correct processing of heterologous structural genes in
XX CC yeast of the genus Hansenula, and so provides high yields of mature
XX CC proteins and facilitates subsequent purification.
XX SQ Sequence 468 BP; 126 A; 110 C; 112 G; 120 T; 0 other;

Query Match 74.7%; Score 28.4; DB 16; Length 468;
Best Local Similarity 84.2%; Pred. No. 0.016; Mismatches 6; Indels 0; Gaps 0;
Matches 32; Conservative 0;

QY 1 ggctctagaggcgagaaataaatcgcttactggcgga 38
      ||||| ||||| ||||| ||||| ||||| |||||
Db 337 ggctctaaaggcgagaaataaatcgcttactggcgga 374

RESULT 3
AAQ54995
ID AAQ54995 standard; DNA; 219 BP.
XX AC AAQ54995;
XX XX
XX DT 11-JUL-1994 (first entry)
XX DE Desulphatohirudin gene with E. coli-Saccharomyces codon usage.
XX KW Hirudin; HV-1; E. coli; codon usage; desulphatohirudin; HV01; 33Asp;
XX KW 33 Asp; biological activity; thrombosis; thromboembolism; ds.
XX XX
XX OS Hirudo medicinalis.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 14..208
XX FT /*tag= a
XX FT /product= Hirudin_HV-1
XX XX
XX XX EP576792-A.
XX XX
XX XX 05-JAN-1994.
XX XX
XX XX 13-APR-1993; 93EP-0105848.
XX XX

```



```

PR 09-APR-1992; 92HU-0001200.
PA (BIOG ) BIOGAL GYOGYSZERGYAR.
PI Albrecht K, Ambrus G, Barabas E, Barta I, Berk T;
PI Botond K, Egyued C, Erdei J, Kiss P, Klupp T, Koenczoel K;
PI Mate G, Molnari, Moravcsik I, Ott I, Pathy A;
PI Poelya K, Salat J, Tegdes A, Vincze A, Zilahi E;
XX WPI: 1994-009153/02.
DR P-PSDB; AAR47488.
XX
XX Prodn. of recombinant desulphatohirudin HV-1 peptide(s) - using
PT E. coli, Saccharomyces and Streptomyces hosts, for increased
PT yields
XX
XX Disclosure; Page 44; 79pp; English.
XX
XX This sequence represents the hirudin HV-1 gene designed on the
CC basis on E. coli codon usage. This sequence was used in the
CC production of desulphatohirudin HV01 33Asp and desulphatohirudin HV-1
CC 33 Asn. The expressed peptides produced using this sequence have the
CC same biological activity as natural hirudin and can be used in the
CC treatment of thrombosis, thromboembolism, etc. Using naturalised
CC sequences such as this, large amounts of hirudin may be produced
CC stably, with the highest production level achieved being 140-180
CC mg/litre of culture.
XX
XX Sequence 219 BP; 63 A; 46 C; 51 G; 58 T; 1 other;
SQ
Query Match 73.7%; Score 28; DB 15; Length 219;
Best Local Similarity 81.6%; Pred. No. 0.021;
Matches 31; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
OY 1 ggctctagagcgcaaaaaaatcaatgcgttactggcga 38
Db 104 gttctacggtgaaaaaatcaatgtgtcactggcga 141
RESULT 4
AAQ54997
ID AAQ54997 standard; DNA; 567 BP.
XX
XX AAQ54997;
AC
XX
XX 11-JUL-1994 (first entry)
DT
XX
XX Desulphatohirudin gene expression/excretion cassette.
DE
XX
XX Hirudin; HV-1; E. coli; codon usage; desulphatohirudin; HV01; 33Asp;
KW 33 Asn; biological activity; thrombosis; thromboembolism; ss.
XX
XX Hirudo medicinalis.
OS
XX
XX Key Location/Qualifiers
FH -35_signal 43..48
FT /*tag= a
FT -10_signal 68..73
FT /*tag= b
FT misc_signal 83
FT /*tag= c
FT /note= "Transcription initiation site"
FT RBS 260..266
FT /*tag= d
FT /note= "Shine-Dalgarno sequence"
FT sig_peptide 269..352
FT /*tag= e
FT mat_peptide 353..550
FT /*tag= f
FT /product= Hirudin
XX
XX EP576792-A.
PN

```

```

XX
XX 05-JAN-1994.
XX
XX 13-APR-1993; 93EP-0105848.
XX
XX 09-APR-1992; 92HU-0001200.
XX
XX (BIOG ) BIOGAL GYOGYSZERGYAR.
XX
XX Albrecht K, Ambrus G, Barabas E, Barta I, Berk T;
PI Botond K, Egyued C, Erdei J, Kiss P, Klupp T, Koenczoel K;
PI Mate G, Molnari, Moravcsik I, Ott I, Pathy A;
PI Poelya K, Salat J, Tegdes A, Vincze A, Zilahi E;
XX WPI: 1994-009153/02.
DR P-PSDB; AAR47490.
XX
XX Prodn. of recombinant desulphatohirudin HV-1 peptide(s) - using
PT E. coli, Saccharomyces and Streptomyces hosts, for increased
PT yields
XX
XX Disclosure; Page 48; 79pp; English.
XX
XX This sequence represents the an expression/secretion cassette for
CC the expression of hirudin HV-1. This sequence was used in the
CC production of desulphatohirudin HV01 33Asp and desulphatohirudin HV-1
CC 33 Asn. The expressed peptides produced using this sequence have the
CC same biological activity as natural hirudin and can be used in the
CC treatment of thrombosis, thromboembolism, etc. Using naturalised
CC sequences such as this, large amounts of hirudin may be produced
CC stably, with the highest production level achieved being 140-180
CC mg/litre of culture.
XX
XX Sequence 567 BP; 162 A; 123 C; 134 G; 148 T; 0 other;
SQ
Query Match 70.5%; Score 26.8; DB 15; Length 567;
Best Local Similarity 81.6%; Pred. No. 0.08;
Matches 31; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
OY 1 ggctctagagcgcaaaaaaatcaatgcgttactggcga 38
Db 443 ggtctgacggtgaaaaaatcaatgtgtcactggcga 480
RESULT 5
AAQ27601
ID AAQ27601 standard; DNA; 195 BP.
XX
XX AAQ27601;
AC
XX
XX 04-FEB-1993 (first entry)
DT
XX
XX Encodes hirudin derived thrombin inhibitor #2.
DE
XX
XX bleeding time; clotting time; anti-thrombotic;
KW congestive phlebothrombosis.
XX
XX Hirudo medicinalis.
OS
XX
XX JP04197184-A.
PN
XX
XX 16-JUL-1992.
PD
XX
XX 28-NOV-1990; 90JP-0323133.
PF
XX
XX 28-NOV-1990; 90JP-0323133.
PR
XX
XX (MITK ) MITSUI TOATSU CHEM INC.
PA
XX
XX WPI: 1992-288441/35.
DR
XX
XX P-PSDB; AAR26314.
PN

```

PT Thrombin-inhibiting polypeptide(s) more potent than hirudin HV1 -  
 PT produced by transformed Bacillus Subtilis, for treating  
 PT thrombosis

XX Claim 6; Page 1; 24pp; Japanese.

XX This sequence encodes a novel, hirudin derived thrombin inhibitor.

XX Sequence 195 BP; 76 A; 24 C; 42 G; 53 T; 0 other;

Query Match 66.3%; Score 25.2; DB 13; Length 195;  
 Best Local Similarity 78.9%; Pred. No. 0.31;  
 Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgagaaataatcaatgcgttactggcga 38

Db 91 ggatctgatggagagaaataatcaatggttacaggaga 128

# RESULT 6

AAQ1851  
 ID AAQ1851 standard; DNA; 208 BP.

XX AC AAQ1851;

XX 26-MAR-1990 (first entry)

XX Synthetic DNA encoding desulphatohirudin.

XX Desulphatohirudin; synthetic gene; thrombin inhibitor; HV-1 gene;  
 KW plasmid p3010; plasmid p4014.

XX JP01247092-A.

XX 02-OCT-1989.

XX 29-MAR-1988; 88JP-0073200.

XX 29-MAR-1988; 88JP-0073200.

XX (MITK ) MITSUI TOATSU CHEM. INC.

XX WPI; 1989-330037/45.

XX Prepn. of desulphatohirudin protein used as thrombin inhibitor - by  
 PT introducing specific DNA into expression vector, inserting obtd.  
 PT recombinant plasmid into host and incubating.

XX Claim 1; page 485; 6pp; Japanese.

XX The synthetc gene is constructed by dividing the hirudin HV-1 gene into  
 CC 7 parts and ligating to form the gene. It is inserted into pBR322 to  
 CC give plasmid p3010. This is inserted into expression vector pKK223-3 to  
 CC give plasmid p4014. This is used to transform E.coli. It produces  
 CC desulphatohirudin, which is active as a thrombin inhibitor. It allows  
 CC easy mass prodn.

XX Sequence 208 BP; 82 A; 23 C; 43 G; 60 T; 0 other;

Query Match 66.3%; Score 25.2; DB 10; Length 208;  
 Best Local Similarity 78.9%; Pred. No. 0.31;  
 Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgagaaataatcaatgcgttactggcga 38

Db 97 ggatctgatggagagaaataatcaatggttacaggaga 134

# RESULT 7

AAQ27600

ID AAQ27600 standard; DNA; 276 BP.

XX AAQ27600;

XX 04-FEB-1993 (first entry)

XX Encodes hirudin derived thombin inhibitor #1.

XX bleeding time; clotting time; anti-thrombotic;  
 KW congestive phlebothrombosis.

XX Hirudo medicinalis.

XX JP04197184-A.

XX 16-JUL-1992.

XX 28-NOV-1990; 90JP-0323133.

XX 28-NOV-1990; 90JP-0323133.

XX (MITK ) MITSUI TOATSU CHEM INC.

XX WPI; 1992-388441/35.

XX P-PSDB; AAR26313.

XX Thrombin-inhibiting polypeptide(s) more potent than hirudin HV1 -  
 PT produced by transformed Bacillus Subtilis, for treating  
 PT thrombosis

XX Claim 4; Page 1; 24pp; Japanese.

XX This sequence encodes a novel, hirudin derived thrombin inhibitor.

XX Sequence 276 BP; 91 A; 39 C; 66 G; 80 T; 0 other;

Query Match 66.3%; Score 25.2; DB 13; Length 276;  
 Best Local Similarity 78.9%; Pred. No. 0.33;  
 Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgagaaataatcaatgcgttactggcga 38

Db 172 ggatctgatggagagaaataatcaatggttacaggaga 209

# RESULT 8

AAQ13831

ID AAQ13831 standard; DNA; 279 BP.

XX AC AAQ13831;

XX 09-DEC-1991 (first entry)

XX MSP signal peptide-hirudin gene fusion.

XX Major secretion product; expression cassette; desulfatohirudin; ss.

XX Lactococcus lactis LM0230 (DSM 5805), Hirudo medicinalis.

XX Key Location/Qualifiers

FT sig\_peptide 1..81

FT /\*tag= a

FT /note= "MSP signal peptide"

FT mat\_peptide 82..279

FT /\*tag= b

FT /note= "hirudin structural gene."

XX EP449770-A.

XX 02-OCT-1991.

XX 13-MAR-1991; 91EP-0810167.

90EP-0306248.

XX



XX PN DB4140381-A.  
 XX PD 09-JUN-1993.  
 XX PF 07-DEC-1991; 91DE-4140381.  
 XX PR 07-DEC-1991; 91DE-4140381.  
 XX PA (FARH ) HOECHST AG.  
 XX PI Crause P, Habermann P, Schmid G, Tripiel D, Ulmer W;  
 XX WPI; 1993-189473/24.  
 XX New stable synthetic iso:hirudin(s) useful as medicaments -  
 PT having amino acid exchanges at positions 33,52,53,54 and 55, useful  
 PT in treatment of thrombosis.  
 XX Example 2; Page 4; 12pp; German.  
 XX Example 2 describes the prodn. of hirudin variants having Ala  
 CC as N-terminal. Hir5 and Hir6 are used in the prodn. of variant 13  
 CC (Ala1, Glu33, Glu52, Glu53, Glu55) (AAR37666).  
 CC The new isohirudin allows longer storage, and injectable solns. can  
 CC be produced and stored. There is no increase in antigenicity. The  
 CC N-terminus does not influence stability. Isohirudin may be useful  
 CC in thrombosis therapy.  
 XX Sequence 35 BP; 5 A; 12 C; 6 G; 12 T; 0 other;

Query Match 62.1%; Score 23.6; DB 14; Length 35;  
 Best Local Similarity 86.7%; Pred. No. 1.1;  
 Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 9 aggcgaataaaatcaatcgcttactgcca 38  
 Db 32 AGGTGAAGAAAGACCACTGCTTACTGCGCA 3

RESULT 14  
 AAQ81528  
 ID AAQ81528 standard; DNA; 195 BP.  
 AC AAQ81528;  
 XX 29-SEP-1995 (first entry)  
 DT Leech hirudin mutein 33-Cys HVI.  
 DE Hirudin mutant; leech; anticoagulant; antithrombotic; Cys 33 HVI; ss.  
 XX Synthetic.  
 KW Key Location/Qualifiers  
 XX CDS 1..195  
 XX /\*tag= a  
 XX WO9504823-A.  
 XX 16-FEB-1995.  
 XX 23-JUL-1994; 94WO-EP02438.  
 XX 04-AUG-1993; 93EP-0810552.  
 XX (CIBA ) CIBA GEIGY AG.  
 PA (UPCG-) UPC GEN-PHARMA AG.  
 XX Chang JY, Grossenbacher H, Markl W, Maerkl W;  
 WPI; 1995-090901/12.

DR P-PSDB; AAR69100.  
 XX Conjugate consisting of 2-4 residues of desulphatohirudin muteins  
 PT - having hirudin activity, used in compns. to treat thromboses  
 PT and embolism  
 XX Claim 34; Page 38; 44pp; English.  
 XX The wild-type hirudin gene is contd. in a plasmid, pJDB207/GAPFL-  
 CC YHR, described in EP-A-340 170. It is used as a template in a PCR  
 CC to construct plasmid pDP34/GAPFL-YHR(CYS33), a mutant of the  
 CC recombinant desulphatohirudin derivative 1 (HVI) with Cys replacing  
 CC Asp at AA posn. 33. This results in the expression of  
 CC desulphatohirudin-Cys(33) (Cys(33)HVI). The primers used for the PCR  
 CC mutagenesis are AAQ81523-Q81526, primers 1-4 respectively. Primers 1  
 CC and 4 are the upstream and downstream primers for amplification of  
 CC the template DNA; primers 2 and 3 are the internal primers which are  
 CC partly complementary to each other and both cover the site of  
 CC mutation. The resulting plasmid is referred to as pDP34/GAPFL-YHR  
 CC (CYS33) (see AAQ81528/R69100). An expression cassette where the  
 CC desulphatohirudin mutein or deriv. is AAQ81528 is claimed, and  
 CC a desulphatohirudin HVI wherein Asp33 is replaced by Cys (AAR69100)  
 CC is also claimed.  
 XX Sequence 195 BP; 57 A; 38 C; 47 G; 53 T; 0 other;  
 SQ Query Match 62.1%; Score 23.6; DB 16; Length 195;  
 Best Local Similarity 76.3%; Pred. No. 1.5;  
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 ggcctagagcgcaaaaatcaatcgcttactgcca 38  
 Db 91 ggtcttctggtgaaagaaacaaatggtgtaccggtga 128  
 RESULT 15  
 AAQ70319  
 ID AAQ70319 standard; DNA; 196 BP.  
 AC AAQ70319;  
 XX 02-APR-1991 (first entry)  
 DT Sequence encoding desulphatohirudin variant 1 (HIV).  
 DE Anticoagulant; thrombin inhibitor; ds.  
 KW Key Location/Qualifiers  
 XX CDS 1..196  
 XX /\*tag= a  
 XX EP225633-A.  
 XX 16-JUN-1987.  
 XX 09-DEC-1986; 86EP-0117098.  
 XX 29-MAY-1986; 86GB-0013088.  
 XX 12-DEC-1985; 85GB-0030631.  
 XX (CIBA ) CIBA GEIGY AG.  
 PA (PLAN-) PLANTORGAN WERK HEINRICH.  
 PA (CHRI-) PLANTORGANW CHRISTENSEN.  
 XX Meyhack B, Markl W, Helm J;  
 WPI; 1987-164868/24.  
 DR P-PSDB; AAP70225.  
 XX New DNA constructs and hybrid vectors for transformation of yeast  
 PT etc. - useful for prodn. and secretion of protein with hirudin  
 PT activity for use as thrombin inhibitors.

XX Disclosure; p18; 146pp; English.  
XX  
CC The preferred DNA construct of the invention contains the PHO5  
CC promoter and a DNA segment consisting of the PHO5 signal sequence  
CC upstream of and in reading frame with a DNA sequence coding for  
CC mature desulphatohirudin. The segment is under the transcriptional  
CC control of the PHO5 promoter and the 3' flanking sequence of the  
CC PHO5 gene.  
XX  
SQ Sequence 196 BP; 54 A; 49 C; 52 G; 41 T; 0 other;  
  
Query Match 62.1%; Score 23.6; DB 8; Length 196;  
Best Local Similarity 76.3%; Pred. No. 1.5;  
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
  
Qy 1 ggctctagagcgcaaaaaaatcaatgcgttactggcga 38  
||| || ||||| || ||||| |||  
Db 91 ggtctgacggtgaaaaaacaccagtcggtaccggtga 128

Search completed: July 15, 2002, 22:44:55  
Job time: 10068 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:31:51 ; Search time 174.99 seconds  
(without alignments)  
53.341 Million cell updates/sec

Title: US-10-053-641-7

Perfect score: 38

Sequence: 1 ggctctagagcgcaaaaaaatcaatgcgttactgcga 38

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2.6/ptodata/1/ina/PTUS\_COMB.seq.\*
- 6: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID                  | Description        |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1          | 30    | 78.9        | 59     | 1 US-07-859-453E-5  | Sequence 5, Appli  |
| 2          | 30    | 78.9        | 59     | 1 US-07-859-453E-9  | Sequence 9, Appli  |
| c 3        | 30    | 78.9        | 128    | 1 US-08-262-384A-5  | Sequence 5, Appli  |
| 4          | 30    | 78.9        | 235    | 1 US-08-262-384A-7  | Sequence 7, Appli  |
| c 5        | 30    | 78.9        | 235    | 1 US-08-262-384A-8  | Sequence 8, Appli  |
| 6          | 30    | 78.9        | 726    | 1 US-08-262-384A-9  | Sequence 9, Appli  |
| c 7        | 30    | 78.9        | 726    | 1 US-08-262-384A-10 | Sequence 10, Appli |
| 8          | 28.4  | 74.7        | 59     | 1 US-07-859-453E-11 | Sequence 11, Appli |
| 9          | 28    | 73.7        | 195    | 1 US-08-044-506B-27 | Sequence 27, Appli |
| 10         | 26.8  | 70.5        | 563    | 1 US-08-044-506B-22 | Sequence 22, Appli |
| 11         | 25.8  | 67.9        | 106    | 1 US-08-262-384A-4  | Sequence 4, Appli  |
| 12         | 25.2  | 66.3        | 195    | 6 5166318-14        | Sequence 4, Appli  |
| 13         | 25.2  | 66.3        | 276    | 6 5166318-15        | Patent No. 5166318 |
| 14         | 25.2  | 66.3        | 279    | 6 5166318-16        | Patent No. 5166318 |
| 15         | 25.2  | 66.3        | 526    | 6 5166318-17        | Sequence 3, Appli  |
| 16         | 24.2  | 63.7        | 43     | 1 US-07-985-110-9   | Sequence 9, Appli  |
| 17         | 24.2  | 63.7        | 43     | 1 US-08-099-053-9   | Sequence 9, Appli  |
| 18         | 24.2  | 63.7        | 43     | 1 US-08-452-829-9   | Sequence 9, Appli  |
| c 19       | 23.6  | 62.1        | 35     | 1 US-07-985-110-10  | Sequence 10, Appli |
| c 20       | 23.6  | 62.1        | 35     | 1 US-08-099-053-10  | Sequence 10, Appli |
| c 21       | 23.6  | 62.1        | 35     | 1 US-08-452-829-10  | Sequence 10, Appli |
| 22         | 23.6  | 62.1        | 59     | 1 US-07-859-453E-7  | Sequence 7, Appli  |
| 23         | 23.6  | 62.1        | 67     | 6 5422249-9         | Patent No. 5422249 |
| 24         | 23.6  | 62.1        | 122    | 6 5422249-8         | Patent No. 5422249 |
| 25         | 23.6  | 62.1        | 195    | 1 US-08-406-948A-7  | Sequence 7, Appli  |
| 26         | 23.6  | 62.1        | 212    | 6 5180668-10        | Patent No. 5180668 |
| 27         | 23.6  | 62.1        | 217    | 6 5422249-13        | Patent No. 5422249 |

|      |      |      |      |                     |                    |
|------|------|------|------|---------------------|--------------------|
| c 28 | 23.6 | 62.1 | 227  | 2 US-07-982-064-8   | Sequence 8, Appli  |
| 29   | 23.6 | 62.1 | 8491 | 2 US-08-757-439-1   | Sequence 1, Appli  |
| 30   | 22   | 57.9 | 46   | 2 US-08-861-459-7   | Sequence 7, Appli  |
| 31   | 22   | 57.9 | 198  | 1 US-07-910-528-2   | Sequence 2, Appli  |
| 32   | 22   | 57.9 | 198  | 1 US-08-348-972-2   | Sequence 2, Appli  |
| 33   | 22   | 57.9 | 198  | 2 US-08-861-459-3   | Sequence 3, Appli  |
| 34   | 22   | 57.9 | 1082 | 1 US-08-715-252-1   | Sequence 1, Appli  |
| 35   | 22   | 57.9 | 1082 | 2 US-08-453-051-3   | Sequence 1, Appli  |
| 36   | 22   | 57.9 | 1130 | 1 US-08-406-948A-5  | Sequence 5, Appli  |
| c 37 | 20.6 | 54.2 | 38   | 2 US-08-861-459-9   | Sequence 9, Appli  |
| 38   | 20.6 | 54.2 | 45   | 1 US-08-044-506B-12 | Sequence 12, Appli |
| 39   | 20.4 | 53.7 | 46   | 2 US-08-861-459-4   | Sequence 4, Appli  |
| 40   | 20.4 | 53.7 | 195  | 1 US-08-044-506B-28 | Sequence 28, Appli |
| 41   | 20.4 | 53.7 | 201  | 1 US-07-854-596B-1  | Sequence 1, Appli  |
| 42   | 20.4 | 53.7 | 202  | 1 US-08-044-506B-20 | Sequence 20, Appli |
| c 43 | 20.4 | 53.7 | 206  | 1 US-08-044-506B-21 | Sequence 21, Appli |
| 44   | 20.4 | 53.7 | 223  | 1 US-07-854-596B-7  | Sequence 7, Appli  |
| 45   | 20.4 | 53.7 | 420  | 1 US-07-854-596B-8  | Sequence 8, Appli  |

ALIGNMENTS

RESULT 1  
US-07-859-453E-5  
; Sequence 5, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:  
; APPLICANT: Kurfuerst, Manfred  
; APPLICANT: Ruebsamen, Klaus  
; APPLICANT: Schmied, Bernhard  
; APPLICANT: Koerwer, Wolfgang  
; APPLICANT: Schwenen, Juergen  
; APPLICANT: Hoefken, Hans Wolfgang  
; TITLE OF INVENTION: Hlurudin/polyalkylene glycol  
; NUMBER OF INVENTIONS: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS version 7.0  
; SOFTWARE: Wordperfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/859,453E  
; FILING DATE: 29-MAY-1992  
; CLASSIFICATION: 514  
; CLASSIFICATION: A 61 K 37/64  
; CLASSIFICATION: A 61 K 47/48  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/01998  
; FILING DATE: 22-NOV-1990  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 59 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; US-07-859-453E-5

Query Match 78.9%; Score 30; DB 1; Length 59;  
Best Local Similarity 86.8%; Pred. No. 0.0012;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggctctagagcgcaaaaaaatcaatgcgttactgcga 38  
||||| ||||||||| ||||||||| |||||||||

```
; ;  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
;
```



US-08-262-384A-7

Query Match 78.9%; Score 30; DB 1; Length 235;  
Best Local Similarity 86.8%; Pred. No. 0.0014;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatgcgttactgagca 38  
||||| ||||||||| ||||||||| ||||||||| |||||||||  
DB 124 GGCTCTGACGGCGAAGAAACCAAGTCGCTTACTGGCGA 161

RESULT 5

US-08-262-384A-8/c  
; Sequence 8, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kell & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 235 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-262-384A-8

Query Match 78.9%; Score 30; DB 1; Length 235;  
Best Local Similarity 86.8%; Pred. No. 0.0014;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatgcgttactgagca 38  
||||| ||||||||| ||||||||| ||||||||| |||||||||  
DB 116 GGCTCTGACGGCGAAGAAACCAAGTCGCTTACTGGCGA 79

RESULT 6

US-08-262-384A-9  
; Sequence 9, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kell & Weinkauff

; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 726 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-262-384A-9

Query Match 78.9%; Score 30; DB 1; Length 726;  
Best Local Similarity 86.8%; Pred. No. 0.0017; 5; Indels 0; Gaps 0;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatgcgttactgagca 38  
||||| ||||||||| ||||||||| ||||||||| |||||||||  
DB 616 GGCTCTGACGGCGAAGAAACCAAGTCGCTTACTGGCGA 653

RESULT 7

US-08-262-384A-10/c  
; Sequence 10, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kell & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820

;; FILING DATE: 18-JUN-1992  
;; INFORMATION FOR SEQ ID NO: 10:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 726 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
US-08-262-384A-10

Query Match 78.9%; Score 30; DB 1; Length 726;  
Best Local Similarity 86.8%; Pred. No. 0.0017; 5; Indels 0; Gaps 0;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgaaaaataatcggttactggcga 38  
||||| ||||||| || ||||||| |||||  
Db 111 GGTCTGACGGCGAAAAAACCGAGTGGTACTGGCGA 74

## RESULT 8

US-07-859-453E-11  
; Sequence 11, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:

;; APPLICANT: Kurfuerst, Manfred  
;; APPLICANT: Ruebsamen, Klaus  
;; APPLICANT: Schmied, Bernhard  
;; APPLICANT: Koerwer, Wolfgang  
;; APPLICANT: Schweden, Juergen  
;; APPLICANT: Hoefken, Hans Wolfgang  
;; TITLE OF INVENTION: Hrudin/polyalkylene glycol  
;; TITLE OF INVENTION: conjugates  
;; NUMBER OF SEQUENCES: 28  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Keil & Weinlauf  
;; STREET: 1101 Connecticut Avenue  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20036

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
;; COMPUTER: IBM compatible  
;; OPERATING SYSTEM: MS-DOS version 7.0  
;; SOFTWARE: WordPerfect version 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/859,453E  
;; FILING DATE: 29-MAY-1992  
;; CLASSIFICATION: 514  
;; CLASSIFICATION: A 61 K 37/64  
;; CLASSIFICATION: A 61 K 47/48

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/EP90/01998  
;; FILING DATE: 22-NOV-1990  
;; INFORMATION FOR SEQ ID NO: 11:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 59 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: unknown  
;; TOPOLOGY: unknown  
US-07-859-453E-11

Query Match 74.7%; Score 28.4; DB 1; Length 59;  
Best Local Similarity 84.2%; Pred. No. 0.0051;  
Matches 32; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgaaaaataatcggttactggcga 38  
||||| ||||||| || ||||||| |||||  
Db 16 GGTCTAAAGCGGACGCTACCGAGTGGTACTGGCGA 53

## RESULT 9

US-08-044-506B-27  
; Sequence 27, Application US/08044506B  
; Patent No. 5552299  
; GENERAL INFORMATION:  
;; APPLICANT: OTT, Istvan  
;; APPLICANT: KLUPP, Tibor  
;; APPLICANT: MOLNAR, Istvan  
;; APPLICANT: PATTHY, Andras  
;; APPLICANT: BARTA, Istvan  
;; APPLICANT: BARKO nee TOTH, Zsuzsa  
;; APPLICANT: AMBRUS, Gabor  
;; APPLICANT: SALAT, Janos  
;; APPLICANT: TEGDES, Aniko  
;; APPLICANT: MORAVCSIK, Imre  
;; TITLE OF INVENTION: Plasmids and Process for Producing  
;; TITLE OF INVENTION: Recombinant Desulphatohirudin HV-1 Peptides  
;; NUMBER OF SEQUENCES: 31  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Keil & Weinlauf  
;; STREET: 1101 Connecticut Avenue  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20036

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage  
;; COMPUTER: IBM AT-compatible, 80286 processor  
;; OPERATING SYSTEM: MS-DOS version 5.0  
;; SOFTWARE: WordPerfect version 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/044,506B  
;; FILING DATE: 09-APR-1993  
;; CLASSIFICATION: 435  
;; INFORMATION FOR SEQ ID NO: 27:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 195 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear

US-08-044-506B-27

Query Match 73.7%; Score 28; DB 1; Length 195;  
Best Local Similarity 81.6%; Pred. No. 0.0088;  
Matches 31; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgaaaaataatcggttactggcga 38  
||||| ||||||| || ||||||| |||||  
Db 91 GGTCTACGGTGAAAAAATCAATGTCTACTGGCGA 128

## RESULT 10

US-08-044-506B-22  
; Sequence 22, Application US/08044506B  
; Patent No. 5552299  
; GENERAL INFORMATION:

;; APPLICANT: OTT, Istvan  
;; APPLICANT: KLUPP, Tibor  
;; APPLICANT: MOLNAR, Istvan  
;; APPLICANT: PATTHY, Andras  
;; APPLICANT: BARTA, Istvan  
;; APPLICANT: BARKO nee TOTH, Zsuzsa  
;; APPLICANT: AMBRUS, Gabor  
;; APPLICANT: SALAT, Janos  
;; APPLICANT: TEGDES, Aniko  
;; APPLICANT: MORAVCSIK, Imre  
;; TITLE OF INVENTION: Plasmids and Process for Producing  
;; TITLE OF INVENTION: Recombinant Desulphatohirudin HV-1 Peptides  
;; NUMBER OF SEQUENCES: 31  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Keil & Weinlauf  
;; STREET: 1101 Connecticut Avenue  
;; CITY: Washington



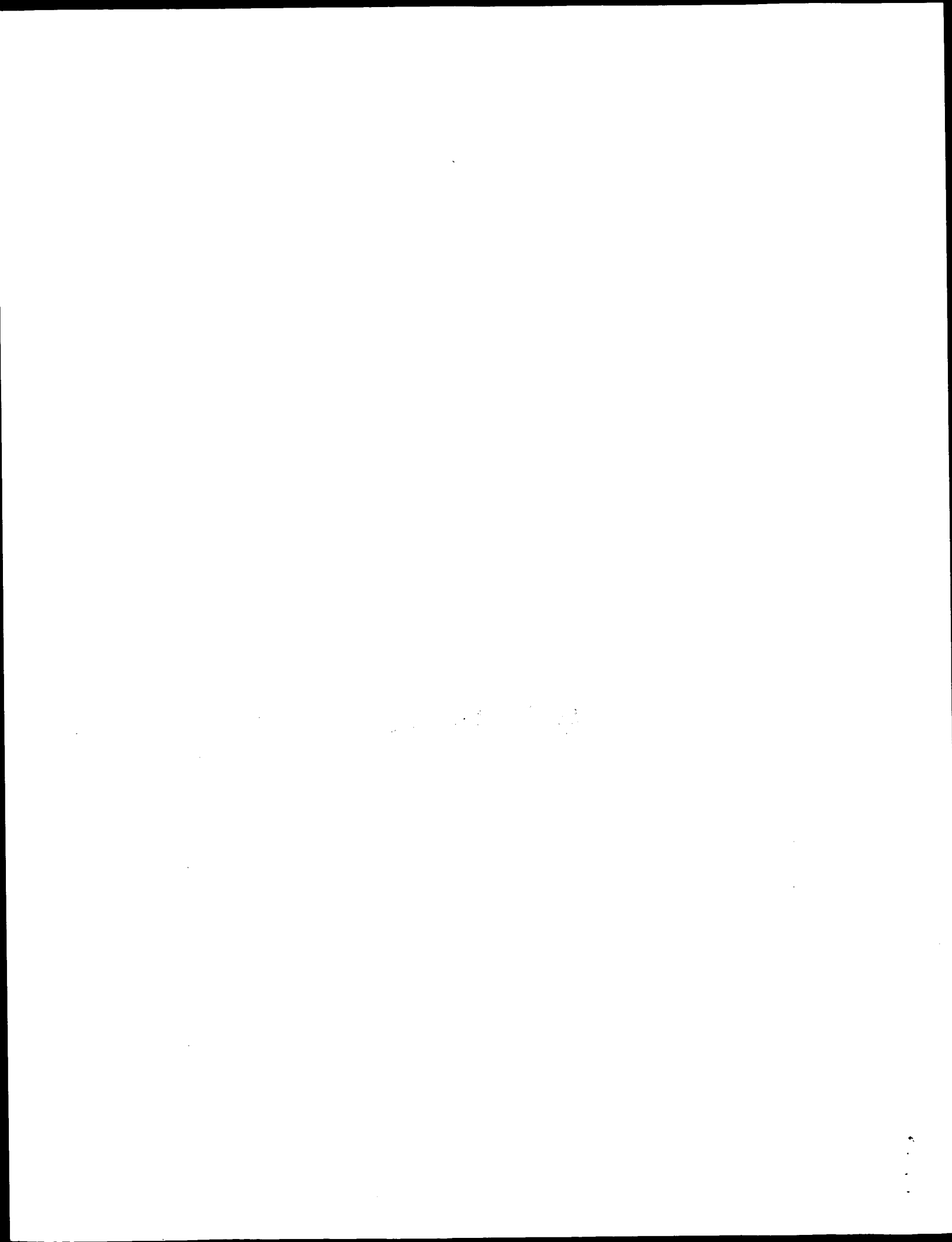
Search completed: July 15, 2002, 22:31:52  
Job time: 18478 sec

```

Query Match      66.3%; Score 25.2; DB 6; Length 526;
Best Local Similarity 78.9%; Pred. No. 0.13;
Matches 30; Conservative 8; Mismatches 0; Gaps 0;
ov      1 gactctadagggcgaataaaatacaatgcgttactgcga 38

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:49:15 ; Search time 6165.88 Seconds  
(without alignments)  
83.181 Million cell updates/sec

Title: US-10-053-641-7

Perfect score: 38

Sequence: 1 ggctctagagcgcaaaaaaatcaatcggttactgcgca 38

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthu:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 23    | 60.5        | 477    | 9  | AW330364           |
| 2          | 21.6  | 56.8        | 473    | 9  | AW634909           |
| 3          | 21.6  | 56.8        | 938    | 12 | CNS03ACU           |
| 4          | 21.4  | 56.3        | 311    | 12 | AL235047 Tetraodon |
| 5          | 21.4  | 56.3        | 451    | 12 | AQ005962 CIT-HSP-2 |
| 6          | 21.4  | 56.3        | 503    | 9  | BH048008 RPI-24-3  |
| 7          | 21.4  | 56.3        | 682    | 10 | AW330322           |
| 8          | 21.4  | 56.3        | 857    | 12 | C96522             |
| 9          | 21.2  | 55.8        | 386    | 10 | BH439433           |
| 10         | 21.2  | 55.8        | 476    | 9  | BI972489           |
| 11         | 21.2  | 55.8        | 559    | 9  | AI900286           |
| 12         | 21.2  | 55.8        | 678    | 12 | AW831943           |
| 13         | 21.2  | 55.3        | 376    | 10 | A2569434           |
| 14         | 21    | 55.3        | 600    | 10 | C64633             |
| 15         | 20.8  | 54.7        | 461    | 12 | BG808735           |
| 16         | 20.8  | 54.7        | 645    | 12 | AQ119701           |
| 17         | 20.8  | 54.7        | 656    | 12 | AQ235385           |
|            |       |             |        |    | AG176441 Pan trogl |

|   |    |      |      |      |    |          |
|---|----|------|------|------|----|----------|
| C | 18 | 20.6 | 54.2 | 313  | 12 | AQ103411 |
|   | 19 | 20.6 | 54.2 | 390  | 9  | AA921897 |
|   | 20 | 20.6 | 54.2 | 428  | 9  | BB638076 |
|   | 21 | 20.6 | 54.2 | 708  | 9  | BB178030 |
| C | 22 | 20.6 | 54.2 | 725  | 9  | BB096074 |
|   | 23 | 20.6 | 54.2 | 727  | 9  | BB096074 |
|   | 24 | 20.6 | 54.2 | 797  | 12 | BB363943 |
|   | 25 | 20.6 | 54.2 | 1020 | 12 | BB363943 |
|   | 26 | 20.4 | 53.7 | 229  | 10 | AQ748677 |
|   | 27 | 20.4 | 53.7 | 344  | 9  | CNS05JFK |
| C | 28 | 20.4 | 53.7 | 404  | 9  | BF891361 |
|   | 29 | 20.4 | 53.7 | 426  | 9  | AA215143 |
| C | 30 | 20.4 | 53.7 | 427  | 10 | AA706088 |
|   | 31 | 20.4 | 53.7 | 449  | 12 | AA706088 |
|   | 32 | 20.4 | 53.7 | 495  | 9  | AW653178 |
|   | 33 | 20.4 | 53.7 | 550  | 9  | R01409   |
| C | 34 | 20.4 | 53.7 | 600  | 10 | AZ847809 |
|   | 35 | 20.4 | 53.7 | 600  | 10 | AI722825 |
|   | 36 | 20.4 | 53.7 | 627  | 9  | AA034037 |
|   | 37 | 20.4 | 53.7 | 686  | 9  | BI985526 |
|   | 38 | 20.4 | 53.7 | 719  | 12 | BI988827 |
|   | 39 | 20.4 | 53.7 | 728  | 9  | BB619126 |
| C | 40 | 20.4 | 53.7 | 806  | 12 | BB642983 |
|   | 41 | 20.4 | 53.7 | 816  | 10 | AG112884 |
|   | 42 | 20.4 | 53.7 | 920  | 9  | AI855745 |
|   | 43 | 20.4 | 53.7 | 1047 | 10 | BH357948 |
| C | 44 | 20.2 | 53.2 | 241  | 9  | BM016367 |
|   | 45 | 20.2 | 53.2 | 389  | 9  | AA826328 |
|   |    |      |      |      |    | BI553059 |
|   |    |      |      |      |    | AV250045 |
|   |    |      |      |      |    | AW488556 |

## ALIGNMENTS

### RESULT 1

AW330364 477 bp mRNA linear EST 21-SEP-2000  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
AW330364  
AW330364  
AW330364.1 GI:6807422  
EST.  
Trypanosoma cruzi.  
Trypanosoma cruzi  
Eukaryota: Euglenozoa: Kinetoplastida: Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477  
/organism="Trypanosoma cruzi"  
/strain="Cl-Brenner"  
/db\_xref="taxon:5693"  
/clone="44h15"  
/clone\_lib="T. cruzi epimastigote"  
/note="cDNA library constructed with oligo dt primed epimastigote mRNA and cloned in pT7c18D phagemid with modified polylinker (Pharmacia)"

BASE COUNT  
ORIGIN

138 a

98 c 105 g 136 t







```

/organism="trypanosoma cruzi"
/strain="Cl-Brenner"
/db_xref="taxon:5693"
/clone="44e7"
/clone_lib="r.cruzi epimastigote normalized cDNA library"
/cell_type="epimastigote"
/notes="cDNA library constructed with oligo dt primed
epimastigote mRNA and cloned in pT7318D phagemid with
modified polylinker (Pharmacia)"
BASE COUNT      148 a 102 c 108 g 145 t
ORIGIN

Query Match      56.3%; Score 21.4; DB 9; Length 503;
Best Local Similarity 80.6%; Pred. NO. 1.2e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcgctta 31
    ||||| ||| ||||| ||| ||| |||
DB 440 GTCCTAAGGGGAAAAAATGACTGCATTA 470

RESULT 7
LOCUS C96522 682 bp mRNA linear EST 27-JUL-1999
DEFINITION C96522 Marchantia polymorpha immature sex organ Marchantia
ACCESSION polymorpha cDNA clone F01W094, mRNA sequence.
VERSION C96522
KEYWORDS EST.
SOURCE C96522.1 GI:4719400
ORGANISM Marchantia polymorpha
          liverwort.
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
          Marchantiophyta; Marchantiopsida; Marchantiales;
          Marchantiaceae; Marchantiaceae; Marchantia.
REFERENCE 1 (bases 1 to 682)
AUTHORS Nagai,J., Yamato,K.T., Sakaida,M., Yoda,H., Fukuzawa,H. and Ohyama
        ,K.
TITLE Expressed sequence tags from immature female sexual organ of a
JOURNAL Liverwort, Marchantia polymorpha
MEDLINE DNA Res. 6 (1), 1-11 (1999)
COMMENT Contact: Nagai J
        Graduate School of Agriculture
        Kyoto University, Division of Applied Life Sciences
        Kitashirakawa-oiwake-cho, Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
        Tel: 81-75-753-6389
        Fax: 81-75-753-6127
        Email: kohnyama@kais.kyoto-u.ac.jp.
FEATURES
        Location/Qualifiers
            1..682
            /organism="Marchantia polymorpha"
            /db_xref="taxon:3197"
            /clone="F01W094"
            /clone_lib="Marchantia polymorpha immature sex organ"
            /sex="female"
            /tissue_type="immature sex organ"
BASE COUNT      177 a 155 c 177 g 173 t
ORIGIN

Query Match      56.3%; Score 21.4; DB 10; Length 682;
Best Local Similarity 80.6%; Pred. NO. 1.2e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcgctta 31
    ||||| ||| ||||| ||| ||| |||
DB 76 GGCTCTCAAGGAGAAAAAATCAAGTGTTA 106

RESULT 8
LOCUS BH439433 857 bp DNA linear GSS 12-DEC-2001

```

```

DEFINITION
BOGUF47TR BOGU Brassica oleracea genomic clone BOGUF47, DNA
sequence.
ACCESSION BH439433
VERSION BH439433.1 GI:17625147
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 857)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOGUF47TF
          Contact: Chris Town
          TIGR
          9712 Medical Center Drive, Rockville, MD 20850, USA.
          Tel: 301-838-3523
          Fax: 301-838-0208
          Email: cdtown@tigr.org
          DNA is from a doubled haploid provided by Tom Osborn.
          Seq primer: TR
          Class: sheared ends.
          Location/Qualifiers
            1..857
            /organism="Brassica oleracea"
            /strain="FO1000DH3"
            /db_xref="taxon:3712"
            /clone="BOGUF47"
            /clone_lib="BOGU"
            /notes="vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
            genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT      216 a 185 c 201 g 255 t
ORIGIN

Query Match      56.3%; Score 21.4; DB 12; Length 857;
Best Local Similarity 80.6%; Pred. NO. 1.1e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 tctagagcgcaaaaaaatcaatcgcttactg 34
    ||||| ||||| ||| ||| ||| |||
DB 742 TCTAGAGCGCAACAAAGTTCAAGGCGCACACTG 772

RESULT 9
LOCUS BI972489 386 bp mRNA linear EST 30-NOV-2001
DEFINITION BI972489 sai78c03.v1 Gm-cl065 Glycine max cDNA clone GENOME SYSTEMS CLONE
ACCESSION ID: Gm-cl065-7014 5', mRNA sequence.
VERSION BI972489
KEYWORDS soybean.
SOURCE Glycine max
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
          Glycine.
REFERENCE 1 (bases 1 to 386)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
        ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
        Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
        ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
        ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
        ,R., Waterston,R. and Willson,R.
        Public Soybean EST Project
        Unpublished (1999)
        Contact: Shoemaker R/Public Soybean EST Project
        Public Soybean EST Project
        Washington University School of Medicine
        4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

```

Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130  
South Memorial Parkway Huntsville, AL 35801 For further information  
call: (800)-533-4363 or contact via email: ccu@resgen.com  
High quality sequence stop: 376.

## FEATURES

source

Location/Qualifiers  
1. 386  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1065-7014"  
/clone\_lib="Gm-c1065"  
/tissue\_type="germinating shoots"  
/lab\_host="DH10B"

/note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2: XhoI; The cDNA library was constructed from mRNA isolated from germinating shoots of the cultivar Williams. The seeds were allowed to germinate for 24 hours prior to being cold stressed for 2 days at 4C. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

BASE COUNT 67 a 142 c 54 g 123 t

## Query Match

Best Local Similarity 55.8%; Score 21.2; DB 10; Length 386;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ctctagagcgcaaaaaaatcaatcg 28  
||||||| |||||||||  
Db 21 CTCTAGAGGTTGAAAAAATCAATCG 46

## RESULT 10

AI900286

LOCUS

DEFINITION AI900286 476 bp mRNA linear EST 30-NOV-2001  
Gm-c1012-1105 5', mRNA sequence.

ACCESSION AI900286

VERSION AI900286.1 GI:5606188

KEYWORDS EST

SOURCE soybean.

ORGANISM

Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

## REFERENCE

AUTHORS

1 (bases 1 to 476)  
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna  
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers  
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk  
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann  
R., Waterston, R. and Wilson, R.

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130

## FEATURES

source

Location/Qualifiers  
1. 476  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1012-1105"  
/clone\_lib="Gm-c1012"  
/tissue\_type="Apical shoot tips, 9-10 day old etiolated  
seedlings"  
/lab\_host="XL10-Gold"

/note="Vector: pBluescript II XR; Site\_1: EcoRI; Site\_2: XhoI; This cDNA library was constructed from mRNA isolated from the apical shoots of 9 to 10 day old etiolated seedlings. The shoot tips including any emerged leaves were harvested for mRNA isolation. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 90 a 154 c 79 g 152 t 1 others

Query Match 55.8%; Score 21.2; DB 9; Length 476;  
Best Local Similarity 88.5%; Pred. No. 1.4e+03;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ctctagagcgcaaaaaaatcaatcg 28  
||||||| |||||||||  
Db 22 CTCTAGAGGTTGAAAAAATCAATCG 47

## RESULT 11

AW831943

LOCUS

DEFINITION AW831943 559 bp mRNA linear EST 03-DEC-2001  
sm18a11.y1 Gm-c1027 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
Gm-c1027-9237 5', mRNA sequence.

ACCESSION AW831943

VERSION AW831943.1 GI:7925917

KEYWORDS EST

SOURCE soybean.

ORGANISM

Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

## REFERENCE

AUTHORS

1 (bases 1 to 559)  
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna  
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers  
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk  
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann  
R., Waterston, R. and Wilson, R.

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130

South Memorial Parkway Huntsville, AL 35801 For further information  
call: (800)-533-4363 or contact via email: ccu@resgen.com  
Seq primer: -40RP from Gibco

High quality sequence stop: 423.  
Location/Qualifiers

South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccu@resgen.com  
Insert Length: 1500 Std Error: 0.00  
High quality sequence stop: 443.

## FEATURES

source

```

1. .559
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl027-9237"
/clone_lib="Gm-cl027"
/tissue_type="cotyledons of 3- and 7-day-old Williams seedlings"
/lab_host="DH10B"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI; This cDNA library was constructed from mRNA isolated from cotyledons of 3- and 7-day-old Williams seedlings which were propagated on paper towels with distilled water. The cotyledons were flash-frozen in liquid nitrogen, then lyophilized for 72 hours. Unequal amounts of mRNA was used for cDNA synthesis. Stratagene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stratagene's first-strand synthesis primer was used. An anchor nucleotide (V-A, C, or G) was added to the 3' end of the primer [GAGAGAGAGAGAGAGAGAGAGAG(T)18] to anchor the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500 bp cutoff, using GibcoBRL Life Technologies' cDNA size fractionation column. The column eluent was then ligated into Stratagene's pBluescript(tm) II XR predigested vector (pBluescript II SK+) that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). 97% of the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=30). This library was constructed by Dr. Paul Keim and Dr. Virginia Corryell."
BASE COUNT      115 a 176 c 90 g 177 t 1 others
ORIGIN

```

```

Query Match      55.8%; Score 21.2; DB 9; Length 559;
Best Local Similarity 88.5%; Pred. No. 1.4e+03;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 3 ctctagagcgcaaaaatcaatgcg 28
|||||  |||||  |||||  |||||  |||||  |||||
Db 9 CTCTAGAGTTGAAAATCAATGCG 34

```

```

RESULT 12
AZ569434 AZ569434 678 bp DNA linear GSS 15-MAY-2001
LOCUS 260PVG03 Pv MBN #30 Plasmodium vivax genomic 3', DNA sequence.
DEFINITION
ACCESSION AZ569434
VERSION AZ569434.1 GI:13979525
KEYWORDS GSS.
SOURCE malaria parasite P. vivax.
ORGANISM Plasmodium vivax
Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
REFERENCE 1 (bases 1 to 678)
AUTHORS Carlton, J.M.-R. and Dame, J.B.
TITLE The Plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL Parasitol. Today 16 (10), 409 (2000)
COMMENT Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine

```

## FEATURES

source

```

1. .678
/organism="Plasmodium vivax"
/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69, 497-598)"
/db_xref="taxon:5855"
/clone_lib="Pv MBN #30"
/dev_stage="asexual blood forms"
/lab_host="Saimiri boliviensis"
/notes="Vector: pBluescript SK(+) vector DNA, phagemid excised from lambda ZAP; Site 1: EcoR V; Site 2: EcoR V; Host leukocytes were extracted from P. vivax infected blood using the following methods: first, infected blood was activated by the addition of 0.5 ml of ADP (40mg/ml) per 10 ml blood. Then blood was passed over a column of acid washed 0.1 mm glass beads, then through a plasmidpur filter, followed by passage through a column of pre-wet Whatman CFF powder (1:2 ratio volume of blood to CFF), and finally centrifuged through a 50% Percoll density cushion. Purified DNA was digested with mung bean nuclease in the presence of 44% formamide at 500c as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). Digested DNA was blunt-ended using T4 DNA polymerase and size fractionated over a Sepharose CL-2B column. Fractions in the size range 500bp-4kb were ligated into the Eco RV site of pBluescript SK(+), and E. coli XL-10 Gold transformed with the ligation mixture."
BASE COUNT      157 a 188 c 156 g 177 t
ORIGIN

```

```

Query Match      55.8%; Score 21.2; DB 12; Length 678;
Best Local Similarity 76.5%; Pred. No. 1.4e+03;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

Qy 1 gctctagcgcaaaaatcaatgcgttactg 34
|||||  |||||  |||||  |||||  |||||  |||||
Db 416 GGGTAAAGGCGAAAAGAGACGCGTCTCTG 449

```

## RESULT 13

```

C64633 C64633 376 bp mRNA linear EST 22-SEP-1997
LOCUS C64633 Yuiji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk357d3 5', mRNA sequence.
ACCESSION C64633
VERSION C64633.1 GI:2423338
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
REFERENCE 1 (bases 1 to 376)
AUTHORS Kohara, Y., Motoshita, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A., and Nishigaki, A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuiji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

```

## FEATURES

```

source
1. .376
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone_lib="y357d3"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT      143 a      78 c      93 g      61 t      1 others
ORIGIN

Query Match      55.3%; Score 21; DB 10; Length 376;
Best Local Similarity 73.0%; Pred. No. 1.6e+03;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 gctctagaggcgagaaaaaatcaatcgcttactggcgga 38
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 319 GATCAACAGCTCAAAAAATAAATGCGTGTCTCGAGA 355

RESULT 14
BG808735
LOCUS
DEFINITION      BG808735 600 bp mRNA linear EST 20-DEC-2001
mRNA sequence.
ACCESSION      BG808735
VERSION      BG808735.1 GI:17955718
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS      Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
White, R.A., Beremand, P.D., Thomas, F.L., Gan, L. and Klein, W.H.
TITLE      Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
JOURNAL      Nucleic Acids Res. 29 (24), 4983-4993 (2001)
COMMENT      Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.

FEATURES
source
1. .600
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"

BASE COUNT      162 a      118 c      123 g      189 t      8 others
ORIGIN

Query Match      55.3%; Score 21; DB 10; Length 600;
Best Local Similarity 71.1%; Pred. No. 1.6e+03;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgagaaaaaatcaatcgcttactggcgga 38
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 450 GGCTCTNTGGGAAAAAATCCATCGGTGACAGAAGA 487

RESULT 15
AQ119701/c
LOCUS
DEFINITION      AQ119701 461 bp DNA linear GSS 22-SEP-1998
HS_2171_B2_H04_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-2171 Col-8 Row-P, DNA sequence.
ACCESSION      AQ119701
VERSION      AQ119701.1 GI:3494468

KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 461)
AUTHORS      Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL      Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE      99380589
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2171 row: P column: 8
Class: BAC ends
High quality sequence stop: 461.
Location/Qualifiers
1. .461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate-2171 Col-8 Row-P"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="organ: sperm; Vector: pBel8BAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT      102 a      76 c      97 g      183 t      3 others
ORIGIN

Query Match      54.7%; Score 20.8; DB 12; Length 461;
Best Local Similarity 75.8%; Pred. No. 1.9e+03;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 6 tagaggcgagaaaaaatcaatcgcttactggcgga 38
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 56 TTGATGCGNAAAAACATCAATAAGATACTGGCAA 24

Search completed: July 15, 2002, 21:49:22
Job time: 18913 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:29:30 ; Search time 2368.24 Seconds  
(without alignments)  
282.762 Million cell updates/sec

Title: US-10-053-641-8  
Perfect score: 32  
Sequence: 1 catgcgcgcctattattgcaggattctt 32

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----|-------------|
| -----      |       |       |        |    |    |             |

|    |      |      |        |    |           |                     |
|----|------|------|--------|----|-----------|---------------------|
| 1  | 21   | 65.6 | 34     | 6  | I82199    | I82199 Sequence 13  |
| 2  | 20.8 | 65.0 | 76     | 6  | A31395    | A31395 DNA fragmen  |
| 3  | 20.8 | 65.0 | 76     | 6  | A31398    | A31398 DNA fragmen  |
| 4  | 20.8 | 65.0 | 128    | 6  | I41185    | I41185 Sequence 5   |
| 5  | 20.8 | 65.0 | 235    | 6  | I41188    | I41188 Sequence 8   |
| 6  | 20   | 62.5 | 197369 | 9  | CNSOIRH6  | AL161662 Human chr  |
| 7  | 19.4 | 60.6 | 68     | 6  | A31391    | A31391 DNA fragmen  |
| 8  | 19.4 | 60.6 | 68     | 6  | A31394    | A31394 DNA fragmen  |
| 9  | 19.4 | 60.6 | 68     | 6  | A31397    | A31397 DNA fragmen  |
| 10 | 19.4 | 60.6 | 68     | 6  | I63550    | I63550 Sequence 13  |
| 11 | 19.4 | 60.6 | 68     | 6  | I63552    | I63552 Sequence 15  |
| 12 | 19.4 | 60.6 | 68     | 6  | I63554    | I63554 Sequence 17  |
| 13 | 19.4 | 60.6 | 77     | 6  | A31392    | A31392 DNA fragmen  |
| 14 | 19.4 | 60.6 | 106    | 6  | I41184    | I41184 Sequence 4   |
| 15 | 19.4 | 60.6 | 226    | 12 | SYNHIR    | M26762 Synthetic h  |
| 16 | 19.4 | 60.6 | 235    | 6  | I41187    | I41187 Sequence 7   |
| 17 | 19.4 | 60.6 | 854    | 33 | AC052752  | AC052752 Giardia i  |
| 18 | 19.4 | 60.6 | 952    | 33 | AC083648  | AC083648 Giardia i  |
| 19 | 19.4 | 60.6 | 1000   | 33 | AC059290  | AC059290 Giardia i  |
| 20 | 19.4 | 60.6 | 1086   | 10 | MM0277752 | AJ277752 Mus muscu  |
| 21 | 19.4 | 60.6 | 51923  | 9  | AL356983  | AL356983 Human DNA  |
| 22 | 19.4 | 60.6 | 215550 | 2  | AL671299  | AL671299 Mus muscu  |
| 23 | 19.4 | 60.6 | 215734 | 2  | AC091784  | AC091784 Mus muscu  |
| 24 | 19.2 | 60.0 | 1778   | 3  | AB018047  | AB018047 Drosophill |
| 25 | 19.2 | 60.0 | 30848  | 3  | AC004307  | AC004307 Drosophill |
| 26 | 19.2 | 60.0 | 46383  | 2  | AC020300  | AC020300 Drosophill |
| 27 | 19.2 | 60.0 | 74717  | 3  | AC004563  | AC004563 Drosophill |
| 28 | 19.2 | 60.0 | 83693  | 8  | AT13138   | AL035524 Arabidops  |
| 29 | 19.2 | 60.0 | 140310 | 2  | AC016009  | AC016009 Homo sapi  |
| 30 | 19.2 | 60.0 | 156617 | 2  | AC099151  | AC099151 Rattus no  |
| 31 | 19.2 | 60.0 | 156799 | 2  | AC079608  | AC079608 Homo sapi  |
| 32 | 19.2 | 60.0 | 179078 | 2  | AC107076  | AC107076 Homo sapi  |
| 33 | 19.2 | 60.0 | 197278 | 9  | AC007367  | AC007367 Homo sapi  |
| 34 | 19.2 | 60.0 | 199749 | 8  | ATCHRIV68 | AL161572 Arabidops  |
| 35 | 19.2 | 60.0 | 259914 | 3  | AE003795  | AE003795 Drosophill |
| 36 | 19.2 | 60.0 | 304290 | 1  | CNSPAX05  | AJ248287 Pyrococcu  |
| 37 | 19.2 | 60.0 | 349980 | 6  | AX041920  | AX041920 Sequence   |
| 38 | 19.2 | 60.0 | 349980 | 6  | AX041921  | AX041921 Sequence   |
| 39 | 19   | 59.4 | 400    | 11 | G31622    | G31622 SHGC-25181   |
| 40 | 19   | 59.4 | 400    | 11 | G34180    | G34180 human STS S  |
| 41 | 19   | 59.4 | 28763  | 2  | AC103135  | AC103135 Rattus no  |
| 42 | 19   | 59.4 | 42458  | 2  | AC103285  | AC103285 Rattus no  |
| 43 | 19   | 59.4 | 63806  | 9  | AL450340  | AL450340 Human DNA  |
| 44 | 19   | 59.4 | 144404 | 8  | AP003210  | AP003210 Oryza sat  |
| 45 | 19   | 59.4 | 145410 | 9  | AC023281  | AC023281 Homo sapi  |

ALIGNMENTS

|            |          |                                     |       |      |        |                 |
|------------|----------|-------------------------------------|-------|------|--------|-----------------|
| RESULT 1   | 182199   | Sequence 13 from patent US 5712114. | 34 bp | DNA  | linear | PAT 10-JUN-1998 |
| LOCUS      | I82199   | Sequence 13 from patent US 5712114. |       |      |        |                 |
| DEFINITION | I82199   | Sequence 13 from patent US 5712114. |       |      |        |                 |
| ACCESSION  | I82199   | Sequence 13 from patent US 5712114. |       |      |        |                 |
| VERSION    | I82199.1 | GI:3210496                          |       |      |        |                 |
| KEYWORDS   |          |                                     |       |      |        |                 |
| SOURCE     |          |                                     |       |      |        |                 |
| ORGANISM   |          |                                     |       |      |        |                 |
| REFERENCE  |          |                                     |       |      |        |                 |
| AUTHORS    |          |                                     |       |      |        |                 |
| TITLE      |          |                                     |       |      |        |                 |
| JOURNAL    |          |                                     |       |      |        |                 |
| FEATURES   |          |                                     |       |      |        |                 |
| Source     |          |                                     |       |      |        |                 |
| BASE COUNT | 5 a      | 8 c                                 | 11 g  | 10 t |        |                 |
| ORIGIN     |          |                                     |       |      |        |                 |

Query Match

65.6% ; Score 21; DB 6; Length 34;

|  |  |                                   |
|--|--|-----------------------------------|
| Best Local Similarity 82.8%; Pred. No. 36;<br>Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  |  |                                   |
| Qy   | 4 gcgccgcccattattgcaggattctt 32<br>                                |                                   |
| Db   | 1 GGGGATCCCTATTACTGCAGGTATTCTT 29<br>                              |                                   |
| RESULT 2   |  |                                   |
| LOCUS  | A31395   | 76 bp DNA linear PAT 03-NOV-1995  |
| DEFINITION   | DNA fragment 3B from patent WO9108229.                             |                                   |
| ACCESSION  | A31395   |                                   |
| VERSION  | A31395.1 GI:1247260  |                                   |
| KEYWORDS   | synthetic construct.   |                                   |
| SOURCE   | synthetic construct  |                                   |
| ORGANISM   | artificial sequence.   |                                   |
| REFERENCE  | 1 (bases 1 to 76)  |                                   |
| AUTHORS  | HIRUDINE POLYALKYLENE GLYCOL CONJUGATES                            |                                   |
| TITLE  | Patent: WO 9108229-A 12 13-JUN-1991;                               |                                   |
| JOURNAL  | Location/Qualifiers  |                                   |
| FEATURES   | source   |                                   |
| source   | 1..76<br>/organism="synthetic construct"<br>/db_xref="taxon:32630" |                                   |
| BASE COUNT   | 12 a 19 c 23 g 22 t  |                                   |
| ORIGIN   |  |                                   |
| Query Match 65.0%; Score 20.8; DB 6; Length 76;<br>Best Local Similarity 91.7%; Pred. No. 48;<br>Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0; |  |                                   |
| Qy   | 9 cgccctattatgcaggattctt 32<br>                                    |                                   |
| Db   | 2 CGACCTATTACTGCAGGTATTCTT 25<br>                                  |                                   |
| RESULT 3   |  |                                   |
| LOCUS  | A31398   | 76 bp DNA linear PAT 03-NOV-1995  |
| DEFINITION   | DNA fragment 3C from patent WO9108229.                             |                                   |
| ACCESSION  | A31398   |                                   |
| VERSION  | A31398.1 GI:1247261  |                                   |
| KEYWORDS   | synthetic construct.   |                                   |
| SOURCE   | synthetic construct  |                                   |
| ORGANISM   | artificial sequence.   |                                   |
| REFERENCE  | 1 (bases 1 to 76)  |                                   |
| AUTHORS  | HIRUDINE POLYALKYLENE GLYCOL CONJUGATES                            |                                   |
| TITLE  | Patent: WO 9108229-A 15 13-JUN-1991;                               |                                   |
| JOURNAL  | Location/Qualifiers  |                                   |
| FEATURES   | source   |                                   |
| source   | 1..76<br>/organism="synthetic construct"<br>/db_xref="taxon:32630" |                                   |
| BASE COUNT   | 11 a 19 c 23 g 23 t  |                                   |
| ORIGIN   |  |                                   |
| Query Match 65.0%; Score 20.8; DB 6; Length 76;<br>Best Local Similarity 91.7%; Pred. No. 48;<br>Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0; |  |                                   |
| Qy   | 9 cgccctattatgcaggattctt 32<br>                                    |                                   |
| Db   | 2 CGACCTATTACTGCAGGTATTCTT 25<br>                                  |                                   |
| RESULT 4   |  |                                   |
| LOCUS  | I41185   | 128 bp DNA linear PAT 13-MAY-1997 |

|  |   |                                      |
|--|---|--------------------------------------|
| Sequence 5 from patent US 5624822.   |   |                                      |
| ACCESSION  | I41185  |                                      |
| VERSION  | I41185.1 GI:2081775   |                                      |
| KEYWORDS   | Unknown.  |                                      |
| SOURCE   | Unknown.  |                                      |
| ORGANISM   | Unclassified.   |                                      |
| REFERENCE  | 1 (bases 1 to 128)  |                                      |
| AUTHORS  | Koerwer, W.   |                                      |
| TITLE  | Hirudin fusion proteins and preparation of hirudin  |                                      |
| JOURNAL  | Patent: US 5624822-A 5 29-APR-1997;   |                                      |
| FEATURES   | Location/Qualifiers   |                                      |
| source   | 1..128<br>/organism="unknown"   |                                      |
| BASE COUNT   | 20 a 33 c 34 g 41 t   |                                      |
| ORIGIN   |   |                                      |
| Query Match 65.0%; Score 20.8; DB 6; Length 128;<br>Best Local Similarity 91.7%; Pred. No. 49;<br>Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  |   |                                      |
| Qy   | 9 cgccctattatgcaggattctt 32<br>   |                                      |
| Db   | 2 CGACCTATTACTGCAGGTATTCTT 25<br>   |                                      |
| RESULT 5   |   |                                      |
| LOCUS  | I41188  | 235 bp DNA linear PAT 13-MAY-1997    |
| DEFINITION   | Sequence 8 from patent US 5624822.  |                                      |
| ACCESSION  | I41188  |                                      |
| VERSION  | I41188.1 GI:2081778   |                                      |
| KEYWORDS   | Unknown.  |                                      |
| SOURCE   | Unknown.  |                                      |
| ORGANISM   | Unclassified.   |                                      |
| REFERENCE  | 1 (bases 1 to 235)  |                                      |
| AUTHORS  | Koerwer, W.   |                                      |
| TITLE  | Hirudin fusion proteins and preparation of hirudin  |                                      |
| JOURNAL  | Patent: US 5624822-A 8 29-APR-1997;   |                                      |
| FEATURES   | Location/Qualifiers   |                                      |
| source   | 1..235<br>/organism="unknown"   |                                      |
| BASE COUNT   | 42 a 58 c 64 g 71 t   |                                      |
| ORIGIN   |   |                                      |
| Query Match 65.0%; Score 20.8; DB 6; Length 235;<br>Best Local Similarity 91.7%; Pred. No. 51;<br>Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  |   |                                      |
| Qy   | 9 cgccctattatgcaggattctt 32<br>   |                                      |
| Db   | 2 CGACCTATTACTGCAGGTATTCTT 25<br>   |                                      |
| RESULT 6   |   |                                      |
| LOCUS  | CNS01RH6  | 197369 bp DNA linear PRI 07-FEB-2002 |
| DEFINITION   | Human chromosome 14 DNA sequence BAC R-439L12 of library RPC1-11 from chromosome 14 of Homo sapiens (Human), complete sequence. |                                      |
| ACCESSION  | AL161662  |                                      |
| VERSION  | AL161662.6 GI:18619912  |                                      |
| KEYWORDS   | HTG.  |                                      |
| SOURCE   | human.  |                                      |
| ORGANISM   | Homo sapiens  |                                      |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.     |                                      |
| AUTHORS  | 1 (bases 1 to 197369)   |                                      |
| Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P., Brottier, P., Cattolico, L., Barbe, V., Pelletier, E., Artiguenave, F., Levy, M., Eckenberg, R., Bruls, T., deBerardinis, V., Cruaud, C., |   |                                      |





```
REFERENCE 1 (bases 1 to 68)
AUTHORS
TITLE HIRUDINE POLYALKYLENE GLYCOL CONJUGATES
JOURNAL Patent: WO 9108229-A 14 13-JUN-1991;
FEATURES Location/Qualifiers
source
BASE COUNT 21 a 21 c 17 g 9 t
ORIGIN

Query Match 60.6%; Score 19.4; DB 6; Length 68;
Best Local Similarity 95.2%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32
||||| |||||||
DB 68 CCTATTACTGCAGGTATTCTT 48
/organism="synthetic construct"
/db_xref="taxon:32630"

RESULT 12
I63554/c 68 bp DNA linear PAT 07-OCT-1997
LOCUS
DEFINITION Sequence 17 from patent US 5663141.
ACCESSION I63554
VERSION I63554.1 GI:2481127
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68)
AUTHORS Kurfuerst,M., Ruebsamen,K., Schmied,B., Koerwer,W., Schweden,J. and
Hoeffken,H.Wolfgang.
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins
JOURNAL Patent: US 5663141-A 17 02-SEP-1997;
FEATURES Location/Qualifiers
source
BASE COUNT 21 a 21 c 17 g 9 t
ORIGIN

Query Match 60.6%; Score 19.4; DB 6; Length 68;
Best Local Similarity 95.2%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32
||||| |||||||
DB 68 CCTATTACTGCAGGTATTCTT 48
/organism="unknown"

RESULT 13
A31392 77 bp DNA linear PAT 03-NOV-1995
LOCUS
DEFINITION DNA fragment 3A from patent WO9108229.
ACCESSION A31392
VERSION A31392.1 GI:1247259
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 77)
AUTHORS
TITLE HIRUDINE POLYALKYLENE GLYCOL CONJUGATES
JOURNAL Patent: WO 9108229-A 9 13-JUN-1991;
FEATURES Location/Qualifiers
source
BASE COUNT 11 a 18 c 23 g 25 t
ORIGIN

Query Match 60.6%; Score 19.4; DB 6; Length 77;
Best Local Similarity 95.2%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32
||||| |||||||
DB 5 CCTATTACTGCAGGTATTCTT 25
/organism="synthetic construct"
/db_xref="taxon:32630"

RESULT 14
I63552/c 68 bp DNA linear PAT 07-OCT-1997
LOCUS
DEFINITION Sequence 15 from patent US 5663141.
ACCESSION I63552
VERSION I63552.1 GI:2481125
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68)
AUTHORS Kurfuerst,M., Ruebsamen,K., Schmied,B., Koerwer,W., Schweden,J. and
Hoeffken,H.Wolfgang.
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins
JOURNAL Patent: US 5663141-A 15 02-SEP-1997;
FEATURES Location/Qualifiers
source
BASE COUNT 20 a 21 c 17 g 10 t
ORIGIN
```

I41184/c  
 LOCUS I41184  
 DEFINITION Sequence 4 from patent US 5624822.  
 ACCESSION I41184  
 VERSION I41184.1 GI:2081774  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE Unclassified.  
 1 (bases 1 to 106)  
 AUTHORS Koerwer, W.  
 TITLE Hirudin fusion proteins and preparation of hirudin  
 JOURNAL Patent: US 5624822-A 4 29-APR-1997;  
 FEATURES  
 source  
 1..106  
 /organism="unknown"  
 BASE COUNT 36 a 28 c 28 g 14 t  
 ORIGIN

Query Match 60.6%; Score 19.4; DB 6; Length 106;  
 Best Local Similarity 95.2%; Pred. No. 2.3e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 12 cctattattgcagggtattctt 32  
 ||||| ||||| ||||| |||||  
 Db 106 CCTATTACTGCAGGTATTCTT 86

RESULT 15  
 SYNHIR/c  
 LOCUS SYNHIR 226 bp DNA linear SYN 27-APR-1993  
 DEFINITION Synthetic hirudin gene, complete cds.  
 ACCESSION M26762  
 VERSION M26762.1 GI:208478  
 KEYWORDS hirudin; proteinase inhibitor; thrombin inhibitor.  
 SOURCE Synthetic DNA.  
 ORGANISM  
 synthetic construct  
 artificial sequence.  
 1 (bases 1 to 226)  
 Bergmann, C., Dödt, J., Koehler, S., Fink, E. and Gassen, H.G.  
 Chemical synthesis and expression of a gene coding for hirudin, the  
 thrombin-specific inhibitor from the leech *Hirudo medicinalis*  
 Biol. Chem. Hoppe-Seyler 367, 731-740 (1986)  
 87026239  
 Location/Qualifiers  
 1..226  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 22..222  
 /note="hirudin"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AA072772.1"  
 /db\_xref="GI:208479"  
 /translation="MVYTDCTSGQNLCLEGSNVGQGNKCILGSDGKNOCVTGE  
 GTPAPQSHNDGDFEIPPEYLQ"  
 BASE COUNT 63 a 56 c 61 g 46 t  
 ORIGIN

CDS  
 1..226  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 22..222  
 /note="hirudin"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AA072772.1"  
 /db\_xref="GI:208479"  
 /translation="MVYTDCTSGQNLCLEGSNVGQGNKCILGSDGKNOCVTGE  
 GTPAPQSHNDGDFEIPPEYLQ"  
 BASE COUNT 63 a 56 c 61 g 46 t  
 ORIGIN

Query Match 60.6%; Score 19.4; DB 12; Length 226;  
 Best Local Similarity 95.2%; Pred. No. 2.4e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 12 cctattattgcagggtattctt 32  
 ||||| ||||| ||||| |||||  
 Db 226 CCTATTACTGCAGGTATTCTT 206

Search completed: July 15, 2002, 22:29:39  
 Job time: 19025 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:55 ; Search time 758.37 Seconds  
(without alignments)  
72.447 Million cell updates/sec

Title: US-10-053-641-8

Perfect score: 32

Sequence: 1 catcgccgcctatttcaggtattctt 32

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: N\_Geneseq\_032802.\*

```

1: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-nbml/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID          | Description         |
|------------|-------|-------------|--------|-------------|---------------------|
| 1          | 21    | 65.6        | 34     | 18 AAT49200 | Hirudin gene linke  |
| C 2        | 19.2  | 60.0        | 2085   | 23 ABL17165 | Drosophila melanog  |
| C 3        | 19.2  | 60.0        | 4466   | 23 ABL17164 | Drosophila melanog  |
| C 4        | 19.2  | 60.0        | 29379  | 23 AAS59510 | Propionibacterium   |
| 5          | 19.2  | 60.0        | 349980 | 22 AAH41225 | Pyrococcus abyssii  |
| 6          | 19.2  | 60.0        | 349980 | 22 AAH41226 | Pyrococcus abyssii  |
| C 7        | 19    | 59.4        | 1539   | 21 AA253344 | Neisseria gonorrhoe |
| 8          | 18.8  | 58.8        | 45     | 16 AAQ88316 | Hirudin gene (codo  |
| 9          | 18.8  | 58.8        | 45     | 16 AAQ85843 | Primer to amplify   |

|      |      |      |       |             |                    |
|------|------|------|-------|-------------|--------------------|
| C 10 | 18.8 | 58.8 | 420   | 12 AAQ12155 | Factor Xa-cleavabl |
| C 11 | 18.8 | 58.8 | 468   | 16 AAQ3244  | Fusion construct o |
| 12   | 18.8 | 58.8 | 737   | 22 AAH05529 | Human cDNA clone ( |
| 13   | 18.8 | 58.8 | 877   | 20 AAX40024 | Prostate cancer as |
| C 14 | 18.8 | 58.8 | 1305  | 23 AAS67097 | DNA encoding novel |
| C 15 | 18.8 | 58.8 | 1467  | 12 AAQ12490 | Factor Xa-cleavabl |
| C 16 | 18.4 | 57.5 | 726   | 12 AAQ12379 | Hirudin peptide/Pr |
| C 17 | 18.4 | 57.5 | 2438  | 23 ABL23342 | Drosophila melanog |
| C 18 | 18.2 | 56.9 | 743   | 22 AA195116 | Human neuroblastom |
| C 19 | 18.2 | 56.9 | 3203  | 11 AAQ06760 | Sequence of beta-g |
| C 20 | 18.2 | 56.9 | 3900  | 23 ABL08940 | Drosophila melanog |
| C 21 | 18.2 | 56.9 | 4324  | 23 ABL16366 | Drosophila melanog |
| C 22 | 18.2 | 56.9 | 5059  | 21 AA236227 | DNA encoding a per |
| C 23 | 18   | 56.2 | 37    | 19 AAV61849 | A. contortrix prot |
| C 24 | 18   | 56.2 | 224   | 7 AAN60746  | DNA encoding hirud |
| 25   | 18   | 56.2 | 850   | 20 AAX40022 | Prostate cancer as |
| 26   | 18   | 56.2 | 2076  | 24 AAS02599 | cDNA sequence #386 |
| 27   | 18   | 56.2 | 2215  | 20 AAX40015 | Prostate cancer as |
| 28   | 18   | 56.2 | 7995  | 24 ABK09796 | Human ovarian tumo |
| 29   | 18   | 56.2 | 8265  | 21 AAC75296 | Human ORFX ORF851  |
| 30   | 18   | 56.2 | 8648  | 22 AAL26716 | Human breast cance |
| 31   | 18   | 56.2 | 8678  | 23 ABL27556 | Drosophila melanog |
| 32   | 18   | 56.2 | 15776 | 23 AAS59583 | Propionibacterium  |
| 33   | 18   | 56.2 | 34878 | 22 AAK66167 | Human immune/haema |
| 34   | 18   | 56.2 | 34878 | 22 AAK80088 | Human immune/haema |
| C 35 | 17.8 | 55.6 | 608   | 23 AAS91030 | DNA encoding novel |
| C 36 | 17.8 | 55.6 | 667   | 22 AAF71326 | Corynebacterium gl |
| C 37 | 17.8 | 55.6 | 1326  | 22 AAH68314 | C glutamicum codin |
| C 38 | 17.8 | 55.6 | 1377  | 20 AAX27119 | Calcineurin cataly |
| C 39 | 17.8 | 55.6 | 1634  | 22 AAS57011 | C. trachomatis DNA |
| C 40 | 17.8 | 55.6 | 1767  | 17 AAT27106 | Yeast calcineurin  |
| C 41 | 17.8 | 55.6 | 2895  | 21 AA64753  | C. trachomatis pmp |
| C 42 | 17.8 | 55.6 | 2895  | 22 AAS6256  | Chlamydia trachoma |
| C 43 | 17.8 | 55.6 | 2934  | 21 AA64759  | C. trachomatis pmp |
| C 44 | 17.8 | 55.6 | 2934  | 22 AAS6262  | Chlamydia trachoma |
| C 45 | 17.8 | 55.6 | 2956  | 21 AA247010 | Feline herpes viru |

#### ALIGNMENTS

#### RESULT 1

AAT49200  
ID AAT49200 standard; DNA; 34 BP.  
XX  
XX AAT49200;  
XX  
DT 14-JUL-1997 (first entry)  
XX  
XX Hirudin gene linked to a signal sequence, PCR 3' primer.  
DE  
XX  
XX Secretion; hirudin; H. polymorpha; polymerase chain reaction;  
KW human; ss.  
XX  
XX Synthetic.  
OS  
XX  
XX WO9639529-A1.  
PN  
XX  
PD 12-DEC-1996.  
XX  
XX 06-JUN-1996; 96WO-US09035.  
PF  
XX  
PR 06-JUN-1995; 95US-0466265.  
XX  
XX (BADI ) BASF AG.  
PA  
XX  
XX Ferenz CR, Hammill LD, Mankovich JA;  
XX  
XX WPI; 1997-043151/04.  
DR  
XX  
XX Pre-pro-collagen signal sequence linked to protein of interest -  
PT esp. hirudin, useful for secretion of heterologous proteins from  
PT host

XX Example 2; Page 15; 45pp; English.

XX The present sequence is the PCR 3' primer for the amplification of a

CC hirudin gene linked to a signal sequence. It is complementary to

CC the carboxy terminus of the Hirudin gene. This preprocollagen signal

CC sequence is used for expressing heterologous proteins in host cells. In a

CC preferred example, a protein is expressed in the yeast Hansenula

CC polymorpha using an expression vector containing a hybrid gene comprising

CC a nucleotide sequence of the human preprocollagen alpha 1(I) protein

CC linked to a nucleotide sequence encoding the heterologous protein of

CC interest. A preferred hybrid gene encodes the preprocollagen signal

CC sequence linked to the protein hirudin. The hybrid genes are useful for

CC secretion of heterologous proteins from host cells.

XX Sequence 34 BP; 5 A; 8 C; 11 G; 10 T; 0 other;

XX

Query Match 65.6%; Score 21; DB 18; Length 34;

Best Local Similarity 82.8%; Pred. No. 2.9;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 gggcgccctattattgcaggattctt 32

DB 1 gggggatccctattactgcaggattctt 29

RESULT 2

ABL17165/C

ID ABL17165 standard; DNA; 2085 BP.

XX

AC ABL17165;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 2968.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Claim 1; SEQ ID NO 2968; 2lpp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

Sequence 34 BP; 5 A; 8 C; 11 G; 10 T; 0 other;

XX

Query Match 65.6%; Score 21; DB 18; Length 34;

Best Local Similarity 82.8%; Pred. No. 2.9;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 gggcgccctattattgcaggattctt 32

DB 1 gggggatccctattactgcaggattctt 29

RESULT 2

ABL17165/C

ID ABL17165 standard; DNA; 2085 BP.

XX

AC ABL17165;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 2968.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Claim 1; SEQ ID NO 2968; 2lpp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 2085 BP; 587 A; 542 C; 503 G; 453 T; 0 other;

Query Match 60.0%; Score 19.2; DB 23; Length 2085;

Best Local Similarity 75.0%; Pred. No. 42;

Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 catgcggccgcctattattgcaggattctt 32

DB 1432 CACGCGGACGCTTGTGTGGCAGTTTCTT 1401

RESULT 3

ABL17164/C

ID ABL17164 standard; DNA; 4466 BP.

XX

AC ABL17164;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 2965.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Claim 1; SEQ ID NO 2965; 2lpp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

Sequence 4466 BP; 1312 A; 945 C; 916 G; 1293 T; 0 other;

XX

Query Match 60.0%; Score 19.2; DB 23; Length 4466;

Best Local Similarity 75.0%; Pred. No. 48;

Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 catgcggccgcctattattgcaggattctt 32

DB 2750 CACGCGGACGCTTGTGTGGCAGTTTCTT 2719



```

AAH41226
ID  AAH41226 standard; DNA; 349980 BP.
XX
AC  AAH41226;
XX
XX  29-OCT-2001 (first entry)
XX
DE  Pyrococcus abyssi genomic fragment #5.
XX
XX  Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX
XX  Pyrococcus abyssi.
XX
XX  Key Location/Qualifiers
XX  misc_feature 1..49980
XX      /tag= a
XX      /note= "This sequence overlaps with the 3' end of
XX      AAH41225"
XX  misc_feature 300001..349980
XX      /tag= b
XX      /note= "This sequence overlaps with the 5' end of
XX      AAH41227"
XX
XX  FR2792651-A1.
XX
XX  27-OCT-2000.
XX
XX  21-APR-1999; 99FR-0005034.
XX
XX  21-APR-1999; 99FR-0005034.
XX
XX  (CNRS ) CNRS CENT NAT RECH SCI.
XX  (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
XX  Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX  Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
XX  WPI; 2001-126236/14.
XX
XX  New nucleotide sequences isolated from Pyrococcus abyssi encode
XX  proteins useful in industry -
XX
XX  Claim 1; Page 511-606; 1657pp; French.
XX
XX  The present invention relates to the genomic sequence of Pyrococcus
XX  abyssi and P. abyssi proteins (see AAB96053-AB9842). P. abyssi is a
XX  hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
XX  vents. The present sequence is a fragment of the genomic sequence of P.
XX  abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41225
XX  and the 3' end of this sequence overlaps with the 5' end of AAH41227. The
XX  proteins of the present invention have various potential industrial uses,
XX  since the proteins are stable at very high temperatures, some up to 110
XX  degrees centigrade.
XX  Note: This patent is in the same patent family as WO200065062, which
XX  contains additional sequences as shown in AAB99132-AAB99143,
XX  AAH75903-AAH75920 and AAG66436.
XX
XX  Sequence 349980 BP; 98084 A; 80447 C; 77665 G; 93784 T; 0 other;

Query Match 60.0%; Score 19.2; DB 22; Length 349980;
Best Local Similarity 75.0%; Pred. No. 1.le+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 catgcggccgcctattattgcaggtattctt 32
||||| ||||| ||||| ||||| |||||
Db 34771 catgagccctcccaataatgaaggtctactt 34802

RESULT 7
AAZ53344/C
ID  AAZ53344 standard; DNA; 1539 BP.
XX
XX
AC  AAZ53344;
XX
XX  21-MAR-2000 (first entry)
XX
XX  Neisseria gonorrhoeae ORF 155 partial DNA sequence SEQ ID NO:637.
XX
XX  Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX  antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX  antibacterial; gene therapy; ds.
XX
XX  Neisseria gonorrhoeae.
XX
XX  WO9957280-A2.
XX
XX  11-NOV-1999.
XX
XX  30-APR-1999; 99WO-US09346.
XX
XX  01-MAY-1998; 98US-0083758.
XX  31-JUL-1998; 98US-0094869.
XX  02-SEP-1998; 98US-0098994.
XX  02-SEP-1998; 98US-0099062.
XX  09-OCT-1998; 98US-0103749.
XX  09-OCT-1998; 98US-0103794.
XX  09-OCT-1998; 98US-0103796.
XX  25-FEB-1999; 99US-0121528.
XX
XX  (CHIR ) CHIRON CORP.
XX  (GENO-) INST GENOMIC RES.
XX
XX  Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX  Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX  Tettelin H, Venter JC;
XX
XX  WPI; 2000-062150/05.
XX
XX  P-PSDB; AA74582.
XX
XX  Novel Neisserial polypeptides predicted to be useful antigens for
XX  vaccines and diagnostics -
XX
XX  Claim 7; Page 444; 1453pp; English.
XX
XX  AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
XX  represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX  and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX  PCR primers used in the exemplification of the present invention. The
XX  polypeptides, the polynucleotides, antibodies and compositions of
XX  the invention can be used as vaccines, as diagnostic reagents, and as
XX  immunogenic compositions. The polypeptides can be used in the
XX  manufacture of medicaments for treating or preventing infection due to
XX  Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX  presence of Neisseria bacteria, or to raise antibodies. They may also
XX  be used to screen for agonists or antagonists, which may themselves
XX  have use as antibacterial agents. The polynucleotides of the invention
XX  may also be used in gene therapy protocols.
XX
XX  Sequence 1539 BP; 330 A; 473 C; 428 G; 308 T; 0 other;

Query Match 59.4%; Score 19; DB 21; Length 1539;
Best Local Similarity 81.5%; Pred. No. 49;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 gcggccgcctattattgcaggtattc 30
||||| ||||| ||||| ||||| |||||
Db 669 GCTGCCGCCGATTCTTCGAGGAAGTC 643

RESULT 8
AAQ88316
ID  AAQ88316 standard; DNA; 45 BP.
XX
XX  AAQ88316;
XX
AC  AAQ88316;

```





FT CDS /note= "encodes cleavage site IEGR"  
 FT 208..408  
 FT /\*tag= c  
 FT /label= hirudin gene  
 XX  
 PN W00109125-A.  
 XX  
 PD 27-JUN-1991.  
 XX  
 XX 07-DEC-1990; 90WO-GB01911.  
 XX  
 XX 07-DEC-1990; 90WO-GB01911.  
 PR  
 PR 07-DEC-1989; 89GB-0027722.  
 XX  
 XX (BRBI-) BRIT BIO-TECHN LTD.  
 PA  
 PI Dawson KM, Hunter MG, Czaplowski LG;  
 XX  
 XX WPI; 1991-208151/28.  
 DR P-PSDB; AAR12888.  
 XX  
 XX Fusion protein cleavage by blood clotting enzyme - for prodn. of  
 PT fractions having greater antithrombotic activity for therapy and  
 PT prophylaxis.  
 XX  
 XX Disclosure; Page 78; 115pp; English.  
 PS  
 XX The sequence of the synthetic hirudin HV-1 genes was designed  
 CC based on the published amino acid sequence (Dott J., et al FEBS  
 CC Letters 165 180 (1984)). Unique restriction sites were incorpor-  
 CC ated to facilitate subsequent genetic manipulation. The codons  
 CC selected were those favourable for *S. cerevisiae* or *E. coli*. The  
 CC sequence was divided into 12 oligomers which were synthesised and  
 CC then annealed. The ligation prod. was ligated to HindIII and EcoRI  
 CC treated pUC19 plasmid DNA and the resulting vector used to trans-  
 CC form *E. coli* K12 HB87. Plasmid pUC19 HV-1 was isolated from trans-  
 CC formants and inserted into plasmid pSW6, a shuttle vector, for ex-  
 CC pression. The gene was then used to construct an expression vector  
 CC in which two hirudin genes are linked together via a linker encod-  
 CC ing a cleavage site for factor Xa. The factor Xa is present at the  
 CC site of the target thrombus so the active agents are released  
 CC specifically at the place where clot formation is occurring.  
 CC See also AAQ12153-Q12156, AAQ12158-Q12162 and AAQ12490.  
 XX  
 XX Sequence 420 BP; 124 A; 90 C; 110 G; 96 T; 0 other;

Query Match 58.8%; Score 18.8; DB 12; Length 420;  
 Best Local Similarity 90.9%; Pred. No. 47;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 11 ccctattattgcaggtattctt 32  
 ||||| ||||| ||||| |||||  
 Db 410 CCCTATTACTGCAGATATCTT 389

RESULT 11  
 AAQ93244/c  
 ID AAQ93244 standard; DNA; 468 BP.

XX AC AAQ93244;

XX DT 01-NOV-1995 (first entry)

XX DE Fusion construct of glucoamylase-hirudin DNA.

XX KW expression cassette; recombinant protein; production; Hansenula;  
 XX KW yeast; leader sequence; adaptor; alpha helix; glucoamylase;  
 XX KW secretion; processing; thrombin inhibitor; hirudin; ds.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT primer\_bind complement (1..33)  
 FT /\*tag= a  
 FT /note= "primer AAQ85840 binding site"  
 FT 13..444  
 FT CDS  
 FT /\*tag= b  
 FT /product= glucoamylase-hirudin fusion protein  
 FT 199..223  
 FT primer\_bind  
 FT /\*tag= c  
 FT /note= "primer AAQ85841 binding site"  
 FT complement (220..267)  
 FT /\*tag= d  
 FT /note= "primer AAQ85842 binding site"  
 FT 13..228  
 FT misc\_feature  
 FT /\*tag= e  
 FT /note= "encodes amino acids 1-72 of glucoamylase"  
 FT 241..246  
 FT misc\_feature  
 FT /\*tag= f  
 FT /note= "encodes processor signal"  
 FT 247..444  
 FT misc\_feature  
 FT /\*tag= g  
 FT /note= "encodes hirudin-h120"  
 FT 0..468  
 FT primer\_bind  
 FT /\*tag= h  
 FT /note= "primer AAQ85843 binding site"  
 FT XX

XX DE4329969-A.

XX PN 09-MAR-1995.

XX PD 04-SEP-1993; 93DE-4329969.

XX PF 04-SEP-1993; 93DE-4329969.

XX PR 04-SEP-1993; 93DE-4329969.

XX XX (BADI ) BASF AG.

XX PA (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.

XX PI Bollschweiler C, Janowicz ZA, Piontek M, Schweden J;

XX PI Strasserawm, Weydemann U;

XX XX WPI; 1995-107810/15.

XX DR P-PSDB; AAR76951.

XX PT Recombinant protein prodn. in Hansenula yeast - transformed with  
 PT expression cassette contg. leader, adaptor, processing signal and  
 PT gene, provides efficient secretion and correct processing

XX XX Example 1; Fig 1; 10pp; German.

XX PS AAQ93244 is a DNA fusion product resulting from ligation of an

CC EcoRI-PvuI fragment (see AAQ85840-41) and a PvuI-SalI fragment, encoding  
 CC a leader sequence, an adaptor (see AAR71472, amino acids 23-72 of GAM  
 CC (glucamylase from Schwanniomycetes occidentalis, plus -His-Pro-Leu-Gln at  
 CC the C-terminal end), a processor peptide (Lys-Arg) and a structural gene,  
 CC encoding hirudin. AAQ93244 is an example of an expression vector insert  
 CC of the general formula: L-A-P-Gene (L = leader sequence; A = adaptor;  
 CC P = processor; Gene = structural gene). The cassettes ensure efficient  
 CC secretion and correct processing of heterologous structural genes in  
 CC yeast of the genus Hansenula, and so provides high yields of mature  
 CC proteins and facilitates subsequent purification.

XX SQ Sequence 468 BP; 126 A; 110 C; 112 G; 120 T; 0 other;

Query Match 58.8%; Score 18.8; DB 16; Length 468;  
 Best Local Similarity 90.9%; Pred. No. 48;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 11 ccctattattgcaggtattctt 32

Db 449 CTCTATTACTGCAGGTATCTT 428

RESULT 12

AAH05529  
ID AAH05529 standard; cDNA; 737 BP.

XX AC AAH05529;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA clone (5'-primer) SEQ ID NO:2364.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
XX  
PR 27-AUG-1999; 99JP-0300253.  
XX  
PR 11-JAN-2000; 2000JP-0118776.  
XX  
PR 02-MAY-2000; 2000JP-0183767.  
XX  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI: 2001-318749/34.  
XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs.  
XX  
PS Claim 1; SEQ ID 2364; 2537pp + CD ROM; English.  
XX

CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification, where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX

SQ Sequence 737 BP; 179 A; 206 C; 212 G; 136 T; 4 other;

Query Match 58.8%; Score 18.8; DB 22; Length 737;  
Best Local Similarity 76.7%; Pred. No. 53;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 tgcggccgcctattatgcagggtattctt 32  
||| ||||| | ||||| ||| |||  
Db 688 tgagccgcgcgtgctattgcagatgctt 717

RESULT 13

AAH0024  
ID AAX40024 standard; DNA; 877 BP.

XX AC AAX40024;  
XX  
DT 02-JUL-1999 (first entry)  
XX  
DE Prostate cancer associated gene.  
XX

KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.

OS Homo sapiens.

PN WO9904265-A2.

PD 28-JAN-1999.

XX 15-JUL-1998; 98WO-US14679.

XX 22-JUN-1998; 98US-0102322.

XX 17-JUL-1997; 97US-0896164.

XX 10-OCT-1997; 97US-0061599.

XX 10-OCT-1997; 97US-0061765.

XX 10-OCT-1997; 97US-0948705.

XX 11-OCT-1997; 97GB-0021697.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;  
XX Pfeurndschuh M, Sahin U, Scanlan MJ, Stockert E;  
XX Tureci O;

XX WPI: 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -  
XX isolated using sera from cancer patients, used to develop products  
XX for the diagnosis, monitoring or treatment of cancers

XX Claim 67; Page 638-639; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised  
XX by expression of a human cancer associated antigen precursor coded for by  
XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
XX biological sample isolated from a subject with an agent that specifically  
XX binds to the NAM, an expression product or a fragment of an expression  
XX product complexed with an HLA molecule; and (b) determining the  
XX interaction between the agent and the NAM or the expression product as a  
XX determination of the disorder. The products and methods can be used in  
XX the diagnosis, monitoring, research, or treatment of conditions  
XX characterised by the expression of various cancer associated antigens.  
XX The invention provides nucleic acid sequences and encoded polypeptides  
XX which are cancer associated antigen precursors expressed in human breast  
XX cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
XX lung cancer.

SQ Sequence 877 BP; 210 A; 248 C; 250 G; 166 T; 3 other;

Query Match 58.8%; Score 18.8; DB 20; Length 877;  
Best Local Similarity 76.7%; Pred. No. 54;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 tgcggccgcctattatgcagggtattctt 32  
||| ||||| | ||||| ||| |||  
Db 680 tgagccgcgcgtgctattgcagatgctt 709

RESULT 14  
AAS67097/c

ID AAS67097 standard; CDNA; 1305 BP.  
 AC AAS67097;  
 DT 13-FEB-2002 (first entry)  
 DE DNA encoding novel human diagnostic protein #2901.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 XX WO200175067-A2.  
 PN 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 PI  
 XX WPI; 2001-639362/73.  
 DR P-PSDB; AEG02910.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 1; SEQ ID No 2901; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 1305 BP; 345 A; 336 C; 357 G; 267 T; 0 other;  
 SQ

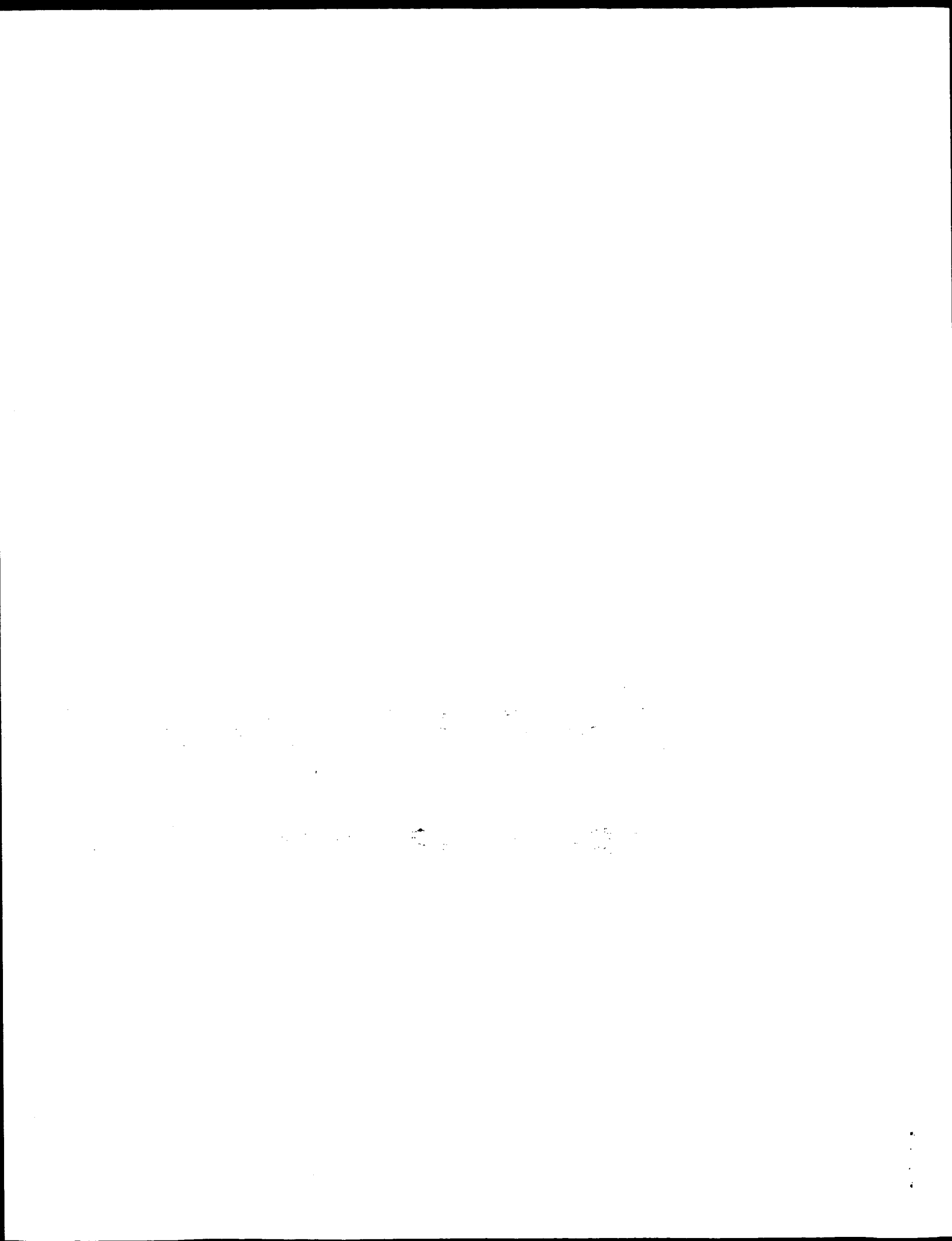
Query Match 58.8%; Score 18.8; DB 23; Length 1305;  
 Best Local Similarity 76.7%; Pred. No. 59;  
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 catgcggcgccctattattgcaggatttc 30  
 ||||| ||| ||| ||| ||| |||  
 Db 1248 CATGGGGTCCACATGTGTGTGATGTTTC 1219

RESULT 15  
 AAO12490/c  
 ID AAO12490 standard; DNA; 1467 BP.

XX AAO12490;  
 AC 17-SEP-1991 (first entry)  
 DT  
 DE Factor Xa-cleavable streptokinase-IEGR-hirudin gene.  
 XX Fusion protein; blood clotting; coagulation; fibrinolysis;  
 KW antithrombotic; thrombolysis; ss.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1455  
 FT /\*tag= a  
 FT mat\_peptide 1..1242  
 FT /\*tag= b  
 FT /label= streptokinase  
 FT misc\_RNA 1243..1254  
 FT /\*tag= c  
 FT /label= linker  
 FT /note= encodes factor Xa cleavage site"  
 FT mat\_peptide 1255..1453  
 FT /\*tag= d  
 FT /label= hirudin HV-1  
 XX  
 PN WO9109125-A.  
 XX 27-JUN-1991.  
 PD  
 XX 07-DEC-1990; 90WO-GB01911.  
 PF  
 XX 07-DEC-1990; 90WO-GB01911.  
 PR 07-DEC-1989; 89GB-0027722.  
 XX (BRBI-) BRIT BIO-TECHN LTD.  
 PA  
 XX Dawson KM, Hunter MG, Czaplewski LG;  
 PI WPI; 1991-208151/28.  
 XX P-PSDB; AARI2522.  
 DR  
 XX Fusion protein cleavage by blood clotting enzyme - for prodn. of  
 PT fractions having greater antithrombotic activity for therapy and  
 PT prophylaxis.  
 XX  
 PS Disclosure; Page 98; 115pp; English.  
 XX The sequence of the synthetic hirudin HV-1 gene was designed  
 CC based on the published amino acid sequence (Dodd J., et al FEBS  
 CC Letters 165 180 (1984)). The sequence of streptokinase was obtd.  
 CC from PCR amplified chromosomal DNA from S. equisimilis ATCC 10009  
 CC or ATCC 9642. The primers used for the PCR were based on the pub-  
 CC lished DNA sequence of S. equisimilis strain H46A (Malke, H., Roe,  
 CC B., and Ferretti, J.J., Gene 34 357-362 [1985]). The two  
 CC streptokinase genes were used to construct an expression vector in which the  
 CC sequence encoding a cleavage site for factor Xa. The factor Xa is  
 CC present at the site of the target thrombus so the active agents are  
 CC released specifically at the place where clot formation is occurring.  
 CC See also AAO12153-Q12156 and AAO12158-Q12162.  
 XX Sequence 1467 BP; 494 A; 317 C; 292 G; 364 T; 0 other;  
 SQ

Query Match 58.8%; Score 18.8; DB 12; Length 1467;  
 Best Local Similarity 90.9%; Pred. No. 60;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 11 ccctattattgcaggatttc 32  
 ||||| ||| ||| ||| ||| |||  
 Db 1457 CCCTATTACTGCAGATATTCTT 1436

Search completed: July 15, 2002, 22:45:19  
Job time: 10092 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:31:52 ; Search time 174.99 seconds  
(without alignments)  
44.918 Million cell updates/sec

Title: US-10-053-641-8

Perfect score: 32

Sequence: 1 catgcggccgcctattattgcaggattctt 32

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description       |
|------------|-------|-------------|--------|----|-------------------|
| 1          | 21    | 65.6        | 34     | 1  | US-08-466-265-13  |
| 2          | 20.8  | 65.0        | 128    | 1  | US-08-262-384A-5  |
| 3          | 20.8  | 65.0        | 235    | 1  | US-08-262-384A-8  |
| C 4        | 19.4  | 60.6        | 68     | 1  | US-07-859-453E-13 |
| C 5        | 19.4  | 60.6        | 68     | 1  | US-07-859-453E-15 |
| C 6        | 19.4  | 60.6        | 68     | 1  | US-07-859-453E-17 |
| C 7        | 19.4  | 60.6        | 105    | 1  | US-08-262-384A-4  |
| C 8        | 19.4  | 60.6        | 235    | 1  | US-08-262-384A-7  |
| C 9        | 18.8  | 58.8        | 31     | 1  | US-07-854-596B-60 |
| 10         | 18.8  | 58.8        | 45     | 1  | US-08-619-598-8   |
| 11         | 18.8  | 58.8        | 45     | 1  | US-08-605-053-5   |
| C 12       | 18.8  | 58.8        | 223    | 1  | US-07-854-596B-7  |
| C 13       | 18.8  | 58.8        | 420    | 1  | US-07-854-596B-8  |
| C 14       | 18.8  | 58.8        | 1467   | 1  | US-07-854-596B-46 |
| C 15       | 18.4  | 57.5        | 726    | 1  | US-08-262-384A-9  |
| 16         | 18.4  | 57.5        | 726    | 1  | US-08-262-384A-10 |
| C 17       | 17.8  | 55.6        | 1767   | 1  | US-08-328-322-20  |
| C 18       | 17.8  | 55.6        | 2353   | 1  | US-08-328-322-11  |
| 19         | 17.6  | 55.0        | 40     | 2  | US-08-839-581A-25 |
| 20         | 17.6  | 55.0        | 40     | 4  | US-09-023-591A-25 |
| 21         | 17.6  | 55.0        | 8491   | 2  | US-08-757-439-1   |
| C 22       | 17.4  | 54.4        | 3157   | 6  | US-08-757-439-1   |
| C 23       | 17.3  | 53.1        | 30     | 1  | US-08-093-741-21  |
| C 24       | 17.3  | 53.1        | 30     | 1  | US-08-720-012-21  |
| C 25       | 17.3  | 53.1        | 30     | 3  | US-08-560-098A-26 |
| C 26       | 17.3  | 53.1        | 30     | 3  | US-08-967-024C-14 |
| 27         | 17    | 53.1        | 50     | 1  | US-08-093-741-24  |

|      |      |      |      |   |                   |                   |
|------|------|------|------|---|-------------------|-------------------|
| 28   | 17   | 53.1 | 50   | 1 | US-08-720-012-24  | Sequence 24, Appl |
| 29   | 17   | 53.1 | 50   | 2 | US-08-560-098A-29 | Sequence 29, Appl |
| 30   | 17   | 53.1 | 50   | 3 | US-08-967-024C-17 | Sequence 17, Appl |
| 31   | 17   | 53.1 | 63   | 3 | US-08-967-024C-20 | Sequence 20, Appl |
| C 32 | 17   | 53.1 | 63   | 3 | US-08-967-024C-21 | Sequence 21, Appl |
| C 33 | 17   | 53.1 | 63   | 3 | US-08-967-024C-22 | Sequence 22, Appl |
| C 34 | 17   | 53.1 | 63   | 3 | US-08-967-024C-23 | Sequence 23, Appl |
| C 35 | 16.8 | 52.5 | 80   | 3 | US-09-282-996-6   | Sequence 6, Appl  |
| C 36 | 16.8 | 52.5 | 201  | 1 | US-07-854-596B-1  | Sequence 1, Appl  |
| C 37 | 16.8 | 52.5 | 282  | 3 | US-08-444-818-120 | Sequence 120, App |
| 38   | 16.8 | 52.5 | 830  | 4 | US-08-913-362-1   | Sequence 1, Appl  |
| C 39 | 16.8 | 52.5 | 1788 | 2 | US-08-722-806A-1  | Sequence 1, Appl  |
| C 40 | 16.8 | 52.5 | 1788 | 4 | US-09-337-028-1   | Sequence 1, Appl  |
| C 41 | 16.8 | 52.5 | 5405 | 3 | US-09-282-996-1   | Sequence 1, Appl  |
| C 42 | 16.8 | 52.5 | 5427 | 3 | US-09-282-996-2   | Sequence 2, Appl  |
| 43   | 16.8 | 52.5 | 6816 | 4 | US-09-404-650-1   | Sequence 1, Appl  |
| 44   | 16.8 | 52.5 | 6855 | 4 | US-09-404-650-3   | Sequence 3, Appl  |
| C 45 | 16.8 | 52.5 | 8987 | 3 | US-08-444-818-137 | Sequence 137, App |

#### ALIGNMENTS

#### RESULT 1

US-08-466-265-13  
Sequence 13, Application US/08466265  
Patent No. 5712114

#### GENERAL INFORMATION:

APPLICANT: Mankovich, John A.  
TITLE OF INVENTION: Compositions for Expression of Proteins in Host Cells Using a Preprocollagen Signal Sequence  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875

#### COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/466,265  
FILING DATE:

#### ATTORNEY/AGENT INFORMATION:

NAME: DeConti, Giulio A., Jr.  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: BBI-030  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941

#### INFORMATION FOR SEQ ID NO:

LENGTH: 34 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: oligonucleotide  
US-08-466-265-13

Query Match 65.6%; Score 21; DB 1; Length 34;

Best Local Similarity 82.8%; Pred. No. 0.42;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 ggcggccgcctattattgcaggattctt 32

Db 1 GGGGGATCCCTACTTACTGCAGGTATTCTT 29





Query Match 60.6%; Score 19.4; DB 1; Length 68;  
Best Local Similarity 95.2%; Pred. No. 2.6;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32  
Db 68 CCTATTACTGCAGGTATTCTT 48

## RESULT 5

US-07-859-453E-15/c  
Sequence 15, Application US/07859453E  
Patent No. 5663141

## GENERAL INFORMATION:

APPLICANT: Kurfuerst, Manfred  
APPLICANT: Ruebsamen, Klaus  
APPLICANT: Schmied, Bernhard  
APPLICANT: Koerwer, Wolfgang  
APPLICANT: Schweden, Juergen  
APPLICANT: Hoeftken, Hans Wolfgang  
TITLE OF INVENTION: Hrudin/polyalkylene glycol  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Kell & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS version 7.0  
SOFTWARE: WordPerfect version 5.1

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/859,453E  
FILING DATE: 29-MAY-1992  
CLASSIFICATION: A 61 K 37/64  
CLASSIFICATION: A 61 K 47/48

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP90/01998

FILING DATE: 22-NOV-1990

INFORMATION FOR SEQ ID NO: 15:

## SEQUENCE CHARACTERISTICS:

LENGTH: 68 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown

US-07-859-453E-15

Query Match 60.6%; Score 19.4; DB 1; Length 68;  
Best Local Similarity 95.2%; Pred. No. 2.6;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32  
Db 68 CCTATTACTGCAGGTATTCTT 48

## RESULT 6

US-07-859-453E-17/c  
Sequence 17, Application US/07859453E  
Patent No. 5663141

## GENERAL INFORMATION:

APPLICANT: Kurfuerst, Manfred  
APPLICANT: Ruebsamen, Klaus  
APPLICANT: Schmied, Bernhard  
APPLICANT: Koerwer, Wolfgang

APPLICANT: Schweden, Juergen  
APPLICANT: Hoeftken, Hans Wolfgang  
TITLE OF INVENTION: Hrudin/polyalkylene glycol  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kell & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS version 7.0  
SOFTWARE: WordPerfect version 5.1

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/859,453E  
FILING DATE: 29-MAY-1992  
CLASSIFICATION: 514  
CLASSIFICATION: A 61 K 37/64  
CLASSIFICATION: A 61 K 47/48

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP90/01998

FILING DATE: 22-NOV-1990

INFORMATION FOR SEQ ID NO: 17:

## SEQUENCE CHARACTERISTICS:

LENGTH: 68 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown

US-07-859-453E-17

Query Match 60.6%; Score 19.4; DB 1; Length 68;  
Best Local Similarity 95.2%; Pred. No. 2.6;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32  
Db 68 CCTATTACTGCAGGTATTCTT 48

## RESULT 7

US-08-262-384A-4/c  
Sequence 4, Application US/08262384A  
Patent No. 5624822

## GENERAL INFORMATION:

APPLICANT: Koerwer, Wolfgang  
TITLE OF INVENTION: The Preparation of Hrudin  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Kell & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM AT-compatible, 80486 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/262,384A  
FILING DATE: 20-JUN-1994  
CLASSIFICATION: 530

CLASSIFICATION: C 12 N 15/62

CLASSIFICATION: C 12 N 15/31

CLASSIFICATION: C 07 K 7/10

CLASSIFICATION: C 12 P 21/02

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP90/02084  
FILING DATE: 04-DEC-1990  
APPLICATION NUMBER: US 07861820  
FILING DATE: 18-JUN-1992  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-262-384A-4

Query Match 60.6%; Score 19.4; DB 1; Length 106;  
Best Local Similarity 95.2%; Pred. No. 2.9;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattatgcagggtattctt 32  
||||| |||||||  
DB 106 CCTATTACTGCAGGTATTCTT 86

## RESULT 8

US-08-262-384A-7/c  
Sequence 7, Application US/08262384A  
Patent No. 5624822  
GENERAL INFORMATION:

APPLICANT: Koerwer, Wolfgang  
TITLE OF INVENTION: The Preparation of Hirudin  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Keil & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM AT-compatible, 80486 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/262,384A  
FILING DATE: 20-JUN-1994  
CLASSIFICATION: 530

CLASSIFICATION: C 12 N 15/62  
CLASSIFICATION: C 12 N 15/31  
CLASSIFICATION: C 07 K 7/10  
CLASSIFICATION: C 12 P 21/02

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP90/02084  
FILING DATE: 04-DEC-1990  
APPLICATION NUMBER: US 07861820

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

## US-08-262-384A-7

Query Match 60.6%; Score 19.4; DB 1; Length 235;  
Best Local Similarity 95.2%; Pred. No. 3.4;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattatgcagggtattctt 32  
||||| |||||||  
DB 235 CCTATTACTGCAGGTATTCTT 215

## RESULT 9

US-07-854-596B-60/c  
Sequence 60, Application US/07854596B  
Patent No. 5434073  
GENERAL INFORMATION:  
APPLICANT: Dawson, Keith M  
APPLICANT: Hunter, Michael G  
APPLICANT: Czaplewski, Lloyd G  
TITLE OF INVENTION: Proteins and nucleic acids  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dr. John J. McDonnell  
STREET: Ten South Wacker Drive, Suite 3000  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/854,596B  
FILING DATE: 03-JUN-1992  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: McDonnell, John J  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 92,337  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
TELEX: 910-221-5317

INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:

NAME/KEY: misc\_feature  
LOCATION: 1..31  
OTHER INFORMATION: /note= "oligonucleotide BB2021"  
US-07-854-596B-60

Query Match 58.8%; Score 18.8; DB 1; Length 31;  
Best Local Similarity 90.9%; Pred. No. 4.2;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 11 cctattatgcagggtattctt 32  
||||| |||||||  
DB 26 CCTATTACTGCAGGTATTCTT 5

## RESULT 10

US-08-619-598-8  
Sequence 8, Application US/08619598  
Patent No. 5672487  
GENERAL INFORMATION:

APPLICANT: SCHWEDEN, Juergen  
APPLICANT: BOLLSCHWEILER, Claus  
APPLICANT: PIONTEK, Michael  
APPLICANT: WEYDEMANN, Ulrike  
APPLICANT: GELLISSEN, Gerd  
TITLE OF INVENTION: THE RECOMBINANT PRODUCTION  
OF PROTEINS IN YEAST  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Keil & Weinkauff  
STREET: 1101 Connecticut Avenue

Query Match  
58.8%; Score 18.8; DB 1; Length 223;

Best Local Similarity 90.9%; Pred. No. 6.3; Indels 2; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

Qy 11 ccctattattgcaggtattctt 32  
||||| ||||| |||||  
Db 219 CCCTATTACTGCAGATATTCTT 198

## RESULT 13

US-07-854-596B-8/c  
; Sequence 8, Application US/07854596B  
; Patent No. 5434073  
; GENERAL INFORMATION:  
; APPLICANT: Dawson, Keith M  
; APPLICANT: Hunter, Michael G  
; APPLICANT: Czaplewski, Lloyd G  
; TITLE OF INVENTION: Proteins and nucleic acids  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dr. John J. McDonnell  
; STREET: Ten South Wacker Drive, Suite 3000  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/854,596B  
; FILING DATE: 03-JUN-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McDonnell, John J  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 92,337  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; TELEX: 910-221-5317  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 420 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..420  
; OTHER INFORMATION: /note= "Factor Xa-cleavable  
; OTHER INFORMATION: Hirudin-IEGR-Hirudin"  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..402  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 1..402  
US-07-854-596B-8

Query Match 58.8%; Score 18.8; DB 1; Length 420;  
Best Local Similarity 90.9%; Pred. No. 7.1;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 11 ccctattattgcaggtattctt 32  
||||| ||||| |||||  
Db 410 CCCTATTACTGCAGATATTCTT 389

## RESULT 14

US-07-854-596B-46/c  
; Sequence 46, Application US/07854596B  
; Patent No. 5434073  
; GENERAL INFORMATION:  
; APPLICANT: Dawson, Keith M  
; APPLICANT: Hunter, Michael G  
; APPLICANT: Czaplewski, Lloyd G  
; TITLE OF INVENTION: Proteins and nucleic acids  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dr. John J. McDonnell  
; STREET: Ten South Wacker Drive, Suite 3000  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/854,596B  
; FILING DATE: 03-JUN-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McDonnell, John J  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 92,337  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; TELEX: 910-221-5317  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1467 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..1467  
; OTHER INFORMATION: /note= "Streptokinase-hirudin  
; OTHER INFORMATION: fusion linked by Factor Xa-cleavable IEGR"  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1449  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 1..1449  
US-07-854-596B-46

Query Match 58.8%; Score 18.8; DB 1; Length 1467;  
Best Local Similarity 90.9%; Pred. No. 9.2; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 11 ccctattattgcaggtattctt 32  
||||| ||||| |||||  
Db 1457 CCCTATTACTGCAGATATTCTT 1436

## RESULT 15

US-08-262-384A-9/c  
; Sequence 9, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinlauf

```

; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262.384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: 530
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 726 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-262-384A-9

```

```

Query Match      57.5%   Score 18.4; DB 1; Length 726;
Best Local Similarity 95.0%   Pred. No. 12;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

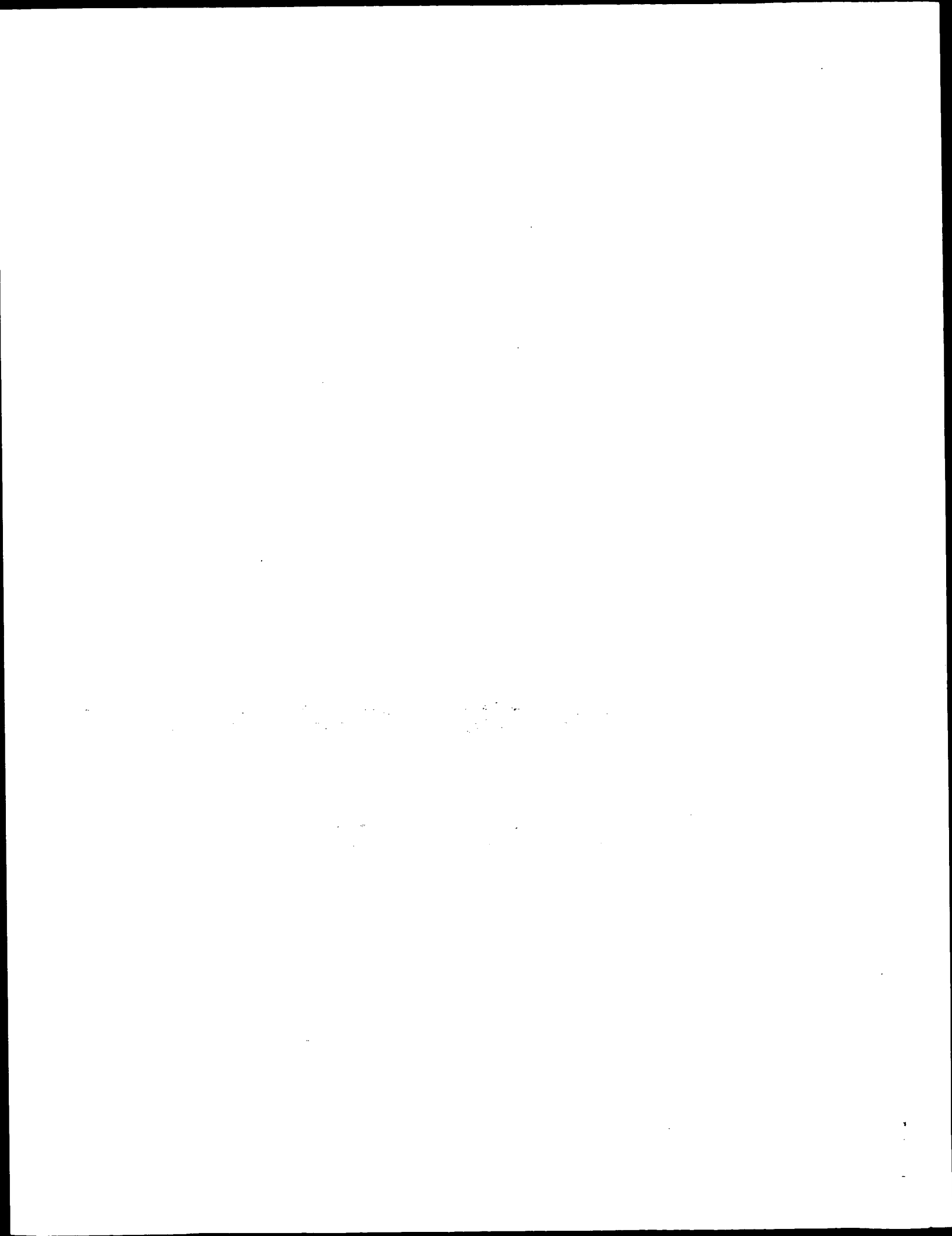
QY 13 ctattattgcaggattcttt 32
Db 726 CTATTACTGCAGGTATTCTT 707

```

```

Search completed: July 15, 2002, 22:31:53
Job time: 18479 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:49:22 ; Search time 6165.88 seconds  
(without alignments)  
70.047 Million cell updates/sec

Title: US-10-053-641-8

Perfect score: 32

Sequence: 1 catcgccgcgcctatttgcaggatttttt 32

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hct:\*
- 9: gb\_estl:\*
- 10: gb\_est2:\*
- 11: gb\_hct:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 21    | 65.6        | 1643   | 10 | BG242109 602354656 |
| 2          | 20.8  | 65.0        | 737    | 10 | BG781087           |
| 3          | 20.4  | 63.7        | 329    | 10 | BG652278           |
| 4          | 20.4  | 63.7        | 509    | 9  | AW317727           |
| 5          | 20.4  | 63.7        | 542    | 10 | BM143360           |
| 6          | 20.4  | 63.7        | 736    | 10 | BE657243           |
| 7          | 20.4  | 63.7        | 1964   | 10 | BF140681           |
| 8          | 20    | 62.5        | 1085   | 10 | BI838672           |
| 9          | 19.8  | 61.9        | 525    | 10 | BM180243           |
| 10         | 19.8  | 61.9        | 593    | 9  | AL628245           |
| 11         | 19.8  | 61.9        | 655    | 9  | AL595404           |
| 12         | 19.8  | 61.9        | 833    | 9  | BE054965           |
| 13         | 19.8  | 61.9        | 1181   | 10 | BE892107           |
| 14         | 19.6  | 61.3        | 952    | 12 | AG098165           |
| 15         | 19.6  | 61.3        | 959    | 10 | BG387896           |
| 16         | 19.6  | 61.3        | 1270   | 12 | AG035742           |
| 17         | 19.4  | 60.6        | 258    | 10 | BG762709 602734617 |

|    |      |      |      |    |          |
|----|------|------|------|----|----------|
| 18 | 19.4 | 60.6 | 438  | 10 | BG359980 |
| 19 | 19.4 | 60.6 | 480  | 9  | AW761595 |
| 20 | 19.4 | 60.6 | 488  | 10 | BI471822 |
| 21 | 19.4 | 60.6 | 646  | 10 | BG599576 |
| 22 | 19.4 | 60.6 | 666  | 10 | W14288   |
| 23 | 19.4 | 60.6 | 685  | 12 | AG074640 |
| 24 | 19.4 | 60.6 | 726  | 12 | AQ048734 |
| 25 | 19.4 | 60.6 | 772  | 12 | AG069748 |
| 26 | 19.4 | 60.6 | 1007 | 10 | BG176216 |
| 27 | 19.2 | 60.0 | 403  | 9  | AI130633 |
| 28 | 19.2 | 60.0 | 409  | 12 | BH618464 |
| 29 | 19.2 | 60.0 | 464  | 12 | B25714   |
| 30 | 19.2 | 60.0 | 521  | 12 | AZ586077 |
| 31 | 19.2 | 60.0 | 544  | 9  | AI313785 |
| 32 | 19.2 | 60.0 | 796  | 10 | BG193247 |
| 33 | 19.2 | 60.0 | 867  | 12 | CNS0101E |
| 34 | 19.2 | 60.0 | 898  | 10 | BF104501 |
| 35 | 19.2 | 60.0 | 925  | 12 | CNS03V97 |
| 36 | 19.2 | 60.0 | 978  | 10 | BF234403 |
| 37 | 19   | 59.4 | 559  | 12 | BH350672 |
| 38 | 19   | 59.4 | 556  | 12 | BH188630 |
| 39 | 19   | 59.4 | 556  | 12 | CNS07SLR |
| 40 | 19   | 59.4 | 669  | 12 | AG074692 |
| 41 | 19   | 59.4 | 692  | 12 | AG125024 |
| 42 | 19   | 59.4 | 723  | 9  | AV704699 |
| 43 | 18.8 | 58.8 | 212  | 9  | AV382007 |
| 44 | 18.8 | 58.8 | 439  | 10 | BG088443 |
| 45 | 18.8 | 58.8 | 483  | 12 | AQ661602 |

#### ALIGNMENTS

#### RESULT 1

BG242109 602354656F1 NCI\_CGAP\_Mam1 Mus musculus cdna clone IMAGE:4483069 5',  
1643 bp mRNA linear EST 13-FEB-2001  
mRNA sequence.  
BG242109 GI:12751924  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 1643)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cyaphs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10320 row: p column: 14  
High quality sequence stop: 34.  
Location/Qualifiers  
1. 1643  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4483069"  
/issue\_type="NCI\_CGAP\_Mam1"  
/tissue\_type="tumor, biopsy sample"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator

#### FEATURES source

```

BASE COUNT      478 a 525 c 283 g 357 t
ORIGIN
    providing samples: Gilbert Smith, NIH"

Query Match      65.6%; Score 21; DB 10; Length 1643;
Best Local Similarity 82.8%; Pred. No. 2e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 ggcggcgccctattattgcaggattctt 32
||||| ||||| ||| ||| ||| |||
Db 25 GCGGCGCCCTATTATTTGTATATATT 53

RESULT 2
BG781087
LOCUS      SEAMC001044 Sea urchin primary mesenchyme cell cDNA library
DEFINITION Strongylocentrotus purpuratus cDNA clone PC_0011_A2_H05_MR 5', mRNA
sequence.
ACCESSION  BG781087
VERSION    BG781087.1 GI:14152100
KEYWORDS   EST.
SOURCE     Strongylocentrotus purpuratus.
ORGANISM   Strongylocentrotus purpuratus.
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinolida; Euechinolida; Echinacea; Echinolida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE  1 (bases 1 to 737)
AUTHORS   Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
            Ettensohn,C.A.
TITLE     A large scale analysis of mRNAs expressed by primary mesenchyme
            cells of the sea urchin embryo
JOURNAL   Development 128 (13), 2615-2627 (2001)
MEDLINE   21384984
COMMENT   Contact: Ettensohn CA
            Dept. Biol. Sci.
            Carnegie Mellon University
            4400 Fifth Avenue, Pittsburgh, PA 15213, USA
            Tel: +1 412 268 5849
            Email: ettensohn@andrew.cmu.edu.

FEATURES
    source
        Location/Qualifiers
            1..737
                /organism="Strongylocentrotus purpuratus"
                /db_xref="taxon:7668"
                /clone="PC_0011_A2_H05_MR"
                /clone_lib="Sea urchin primary mesenchyme cell cDNA
                library"
                /tissue_type="embryo"
                /cell_type="primary mesenchyme cells"
                /lab_host="E.coli"
                /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
                dt priming from poly A+ RNA, directionally cloned"

BASE COUNT      269 a 110 c 122 g 236 t
ORIGIN
    Query Match      65.0%; Score 20.8; DB 10; Length 737;
    Best Local Similarity 78.1%; Pred. No. 2e+02;
    Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 catgcggcgccctattattgcaggattctt 32
||||| ||||| ||||| ||||| |||||
Db 32 CACGCGTCCCATATTATTTGTATATTTCTT 63

RESULT 3
BG652278
LOCUS      sad64h03.y1 Gm-cl051 Glycine max cDNA clone GENOME SYSTEMS CLONE
DEFINITION ID: Gm-cl051-5333 5', mRNA sequence.
ACCESSION  BG652278
VERSION    BG652278.1 GI:13789687

BASE COUNT      329 bp mRNA linear EST 29-NOV-2001
ORIGIN
    Query Match      63.7%; Score 20.4; DB 10; Length 329;
    Best Local Similarity 80.0%; Pred. No. 2.5e+02;
    Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 tgcggcgccctattattgcaggattctt 32
||||| ||||| ||||| ||||| |||||
Db 206 TGCTGCTGCCATATCATTCAGGTAATTTT 235

RESULT 4
AW317727
LOCUS      sg56e07.y1 Gm-cl1007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl1007-445 5', mRNA sequence.
ACCESSION  AW317727
VERSION    AW317727.1 GI:6747271
KEYWORDS   EST.
SOURCE     soybean.
ORGANISM   Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

KEYWORDS
SOURCE
ORGANISM
EST.
    soybean.
    Glycine max
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
    Glycine.
    1 (bases 1 to 329)
    Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
    ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
    Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
    ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
    ,R., Ritter,B., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
    ,R., Waterston,R. and Willson,R.
    Public Soybean EST Project
    Unpublished (1999)
    Contact: Shoemaker R/Public Soybean EST Project
    Public Soybean EST Project
    Washington University School of Medicine
    4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
    Tel: 314 286 1800
    Fax: 314 286 1810
    Email: est@watson.wustl.edu
    This clone is available through: ResGen, Invitrogen Corp. 2130
    South Memorial Parkway Huntsville, AL 35801 For further information
    call: (800) 533-4363 or contact via email: ccu@resgen.com
    High quality sequence stop: 316.
    Location/Qualifiers
        1..329
            /organism="Glycine max"
            /db_xref="taxon:3847"
            /clone="GENOME SYSTEMS CLONE ID: Gm-cl051-5333"
            /clone_lib="Gm-cl051"
            /tissue_type="floral meristematic mRNA"
            /lab_host="DH10B"
            /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
            XhoI; The cDNA library was constructed from floral
            meristematic mRNA provided by Dr. Halina Knap of Clemson
            University. Complementary DNA was synthesized from mRNA
            using a primer consisting of a poly(dT) sequence with a
            XhoI restriction site. EcoRI adapters were ligated to the
            blunt-ended cDNA fragments followed by XhoI digestion. The
            cDNA fragments were directionally cloned into the
            EcoRI-XhoI restriction site of the pBluescript vector. The
            ligated cDNA fragments were transformed into DH10B host
            cells (GibcoBRL). This library was constructed in the
            laboratory of Dr. Randy Shoemaker."
        90 a 59 c 58 g 122 t
    Query Match      63.7%; Score 20.4; DB 10; Length 329;
    Best Local Similarity 80.0%; Pred. No. 2.5e+02;
    Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 tgcggcgccctattattgcaggattctt 32
||||| ||||| ||||| ||||| |||||
Db 206 TGCTGCTGCCATATCATTCAGGTAATTTT 235

RESULT 4
AW317727
LOCUS      sg56e07.y1 Gm-cl1007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl1007-445 5', mRNA sequence.
ACCESSION  AW317727
VERSION    AW317727.1 GI:6747271
KEYWORDS   EST.
SOURCE     soybean.
ORGANISM   Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

```



(bases 1 to 542)  
 hoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V., Khanna  
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,

RESULT 6  
BE657243/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION

|              | 736 bp     | max     | linear | EST 24-MAY-2001       |
|--------------|------------|---------|--------|-----------------------|
| E657243      | max        | Glycine | clone  | Gm-r1070-112 3', mRNA |
| MF70001420C8 | Gm-r1070   |         |        |                       |
| sequence.    |            |         |        |                       |
| E657243      |            |         |        |                       |
| E657243.1    |            |         |        |                       |
| E657243.1    | GI:9983135 |         |        |                       |



CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: InCyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM11566 row: a column: 14  
 High quality sequence stop: 601.  
 Location/Qualifiers  
 1..1085  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5224981"  
 /clone\_lib="NIH\_MGC\_120"  
 /lab\_host="DH10B"  
 /note="Organ: pooled pancreas and spleen; Vector:  
 pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA  
 source anonymous pool of spleen and pancreas from 28 yo  
 male. Library is oligo-dr primed and directionally cloned  
 (EcoRV site is destroyed upon cloning). Average insert  
 size 1.5 kb, insert size range 1-2.5 kb. Library is  
 normalized and enriched for full-length clones and was  
 constructed by C. Gruber (Invitrogen). Research Genetics  
 tracking code 025. Note: this is a NIH\_MGC Library."

#### FEATURES

source

BASE COUNT 332 a 306 c 338 g 109 t  
 ORIGIN

Query Match 62.5%; Score 20; DB 10; Length 1085;  
 Best Local Similarity 82.1%; Pred. No. 4.7e+02;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 ggcggccgcctattatgcagggtattct 31  
 ||||| ||||| || ||||| |||||  
 Db 540 GCGGGCGCCCTCTTTTGCAGCTTTCT 513

#### RESULT

BM180243 9  
 LOCUS BM180243 525 bp mRNA linear EST 07-DEC-2001  
 DEFINITION daJ87b06.x1 NICHD XGC L11 Xenopus laevis cDNA clone IMAGE:5129795  
 3', mRNA sequence.  
 BM180243  
 ACCSSION BM180243.1 GI:17404314  
 VERSION EST.  
 KEYWORDS African clawed frog.  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 Xenopodinae; Xenopus.  
 1 (bases 1 to 525)  
 NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaps.r@mail.nih.gov  
 Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone distribution: Xenopus clones from this library are available  
 through the I.M.A.G.E. Consortium/LLNL at: [infoimage.llnl.gov](http://infoimage.llnl.gov)  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 427.  
 Location/Qualifiers  
 1..525  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="IMAGE:5129795"  
 /clone\_lib="NICHD XGC L11"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dr.

#### FEATURES

source

Query Match 61.9%; Score 19.8; DB 9; Length 593;  
 Best Local Similarity 77.4%; Pred. No. 5e+02;  
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 catcgccgcgcctattatgcagggtattct 31  
 || ||||| ||||| ||||| |||||  
 Db 549 CAAGCGGGGGCCCTACTCTGCAGTCTTCT 579

#### RESULT

AL595404 11  
 LOCUS AL595404 655 bp mRNA linear EST 30-JUL-2001  
 DEFINITION AL595404 XGC-gastrula silurana tropicalis cDNA clone TGas006g19 5',  
 mRNA sequence.

BASE COUNT 170 a 77 c 101 g 177 t  
 ORIGIN

Query Match 61.9%; Score 19.8; DB 10; Length 525;  
 Best Local Similarity 77.4%; Pred. No. 4.8e+02;  
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 atcgccgcgcctattatgcagggtattct 32  
 ||||| || ||||| ||||| |||||  
 Db 97 ATGCGCCCTGTTGTAFTATTCATGCAFTCTT 127

#### RESULT

AL628245 10  
 LOCUS AL628245 593 bp mRNA linear EST 02-NOV-2001  
 DEFINITION AL628245 XGC-gastrula silurana tropicalis cDNA clone TGas009a06 5',  
 mRNA sequence.  
 AL628245  
 ACCSSION AL628245.1 GI:16597728  
 VERSION EST.  
 KEYWORDS western clawed frog.  
 SOURCE  
 ORGANISM  
 Silurana tropicalis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 Xenopodinae; Silurana.  
 1 (bases 1 to 593)  
 Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.  
 Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
 Unpublished (2001)  
 Contact: Huckle E  
 Sanger Centre  
 Hinxtion, Cambridgeshire, CB10 1SA, UK  
 Email: trop@sanger.ac.uk  
 Sanger Xenopus tropicalis EST project 2001  
 TROPICALIS\_SEQUENCE\_ID: TGas009a06.sp6  
 Sequencing primer: SP6  
 This sequence is from a Xenopus Gene Collection (XGC) library  
 constructed by Aaron M. Zorn.  
 Location/Qualifiers  
 1..593  
 /organism="Silurana tropicalis"  
 /db\_xref="taxon:8364"  
 /clone="TGas009a06"  
 /clone\_lib="XGC-gastrula"  
 /dev\_stage="gastrula (stages 10.5-13 mixed)"  
 /lab\_host="Escherichia coli XL1-blue"  
 /note="Vector: pCSI07; Site\_1: EcoRI; Site\_2: NotI; CDNA  
 was oligo dr primed from 5ug of poly A+ RNA from stages  
 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated  
 into pCSI07 with EcoRI at the 5' end and NotI at the 3'  
 end."

#### FEATURES

source

BASE COUNT 143 a 135 c 169 g 146 t  
 ORIGIN

Query Match 61.9%; Score 19.8; DB 9; Length 593;  
 Best Local Similarity 77.4%; Pred. No. 5e+02;  
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 catcgccgcgcctattatgcagggtattct 31  
 || ||||| ||||| ||||| |||||  
 Db 549 CAAGCGGGGGCCCTACTCTGCAGTCTTCT 579

```

ACCESSION      AL595404
VERSION        AL595404.1  GI:15007479
KEYWORDS
SOURCE         western clawed frog.
ORGANISM       Silurana tropicalis
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
               Xenopodinae; Silurana.
REFERENCE      1 (bases 1 to 655)
AUTHORS        Huckle, E., Taylor, R., McMurray, A., Ashurst, J.L., Zorn, A.M. and
               Rogers, J.
TITLE          Sanger Xenopus tropicalis EST project 2001
JOURNAL        Unpublished (2001)
COMMENT        Contact: Huckle E
               Sanger Centre
               Hinxton, Cambridgeshire, CB10 1SA, UK
               Email: trop@sanger.ac.uk
               Sanger Xenopus tropicalis EST project 2001
               TROPICALIS_SEQUENCE_ID: Tgas006g19.sp6
               Sequencing primer: SP6
               This sequence is from a Xenopus Gene Collection (XGC) library
               constructed by Aaron M. Zorn.
               Location/Qualifiers
FEATURES       source
               1..655
               /organism="Silurana tropicalis"
               /db_xref="taxon:8364"
               /clone="Tgas006g19"
               /clone_lib="XGC-gastrula"
               /dev_stage="gastrula (stages 10.5-13 mixed)"
               /lab_host="Escherichia coli XL1-blue"
               /note="vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
               was oligo dt primed from 5ug of poly A+ RNA from stages
               10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
               into pCS107 with EcoRI at the 5' end and NotI at the 3'
               end."
BASE COUNT     169 a 148 c 180 g 158 t
ORIGIN
Query Match   61.9%; Score 19.8; DB 9; Length 655;
Best Local Similarity 77.4%; Pred. No. 5.1e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 catgcggccgcctatttcagcaggtattct 31
||||| ||||||| ||||||| ||||||| |||||||
DB 579 CAAGCGGGGGCCCTATTCTCGAGTCTCTCT 609

RESULT 12
LOCUS          BE054965
DEFINITION     GA_Ea0032D15f Gossypium arboreum 7-10 dpa fiber library Gossypium
               arboreum cDNA clone GA_Ea0032D15f, mRNA sequence.
ACCESSION      BE054965
VERSION        BE054965.2  GI:13245904
KEYWORDS
SOURCE         Gossypium arboreum.
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE      1 (bases 1 to 833)
AUTHORS        Wing, R.A., Frisch, D., Yu, V., Main, D., Rambo, T., Simmons, J., Henry
               D., Wood, T.C., Leslie, A. and Wilkins, T.A.
TITLE          An integrated analysis of the genetics, development, and evolution
               of the cotton fiber
JOURNAL        Unpublished (2000)
COMMENT        On Jun 8, 2000 this sequence version replaced gi:8382022.
               Contact: Wing RA
               Clemson University Genomics Institute
               100 Jordan Hall, Clemson, SC 29634, USA
               Tel.: 864 656 7288

```

```

Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
High quality sequence stop: 765.
FEATURES       source
               1..833
               /organism="Gossypium arboreum"
               /strain="AKA"
               /cultivar="8400"
               /db_xref="taxon:29729"
               /clone="GA_Ea0032D15f"
               /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
               /tissue_type="Fibers isolated from bolls harvested 7-10
               dpa"
               /lab_host="E. coli"
               /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT     231 a 189 c 175 g 238 t
ORIGIN
Query Match   61.9%; Score 19.8; DB 9; Length 833;
Best Local Similarity 77.4%; Pred. No. 5.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 atgcggccgcctatttcagcaggtattctt 32
||||| ||||||| ||||||| ||||||| |||||||
DB 702 ATGGGGTCTCCCTATTATTATGGCTATCTT 732

RESULT 13
LOCUS          BE892107
DEFINITION     BE892107 1181 bp mRNA linear EST 20-OCT-2000
               601434612F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3919898 5',
               mRNA sequence.
ACCESSION      BE892107
VERSION        BE892107.1  GI:10352103
KEYWORDS       EST.
SOURCE         human.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1181)
AUTHORS        NIH-MGC http://mgi.nci.nih.gov/.
               National Institutes of Health, Mammalian Gene Collection (MGC)
               Unpublished (1999)
               Contact: Robert Strausberg, Ph.D.
               Email: cgapbs@email.nih.gov
               Tissue Procurement: ATCC/DCTD/DTP
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM9750 row: g column: 03
               High quality sequence stop: 255.
FEATURES       source
               1..1181
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:3919898"
               /clone_lib="NIH_MGC_72"
               /tissue_type="melanotic melanoma"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: skin; Vector: pCMV-Sport6; Site_1: NotI;
               Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
               Average insert size 2 kb. Library constructed by Life
               Technologies."
BASE COUNT     426 a 300 c 294 g 161 t
ORIGIN
Query Match   61.9%; Score 19.8; DB 10; Length 1181;

```

Best Local Similarity 77.4%; Pred. No. 5.7e+02;  
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 catcgccgcctattattgcaggtattct 31  
||||| ||||| ||||| |||||  
Db 1015 CCGTGGCTGCTATTGTTGCACGTATTCT 985

RESULT 14

AG098165  
LOCUS Pan troglodytes DNA, clone: PTB-100A23.F, genomic survey sequence.  
DEFINITION AG098165  
VERSION AG098165.1 GI:16718682  
KEYWORDS GSS; GSS (genome survey sequence).  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male BAC Library clone:PTB-100A23.F.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1 (sites) Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE

BAC end sequences of library PTB

JOURNAL

2 (bases 1 to 952) Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpanzee@gsr.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/; Tel: 81-45-503-9111, Fax: 81-45-503-9170).

AUTHORS

Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.  
PRIMERS  
Sequencing: -21M13  
LIBRARY  
Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI  
Location/Qualifiers  
1. .952  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-100A23.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"  
BASE COUNT 131 a 346 c 214 g 253 t 8 others  
ORIGIN

Query Match 61.3%; Score 19.6; DB 12; Length 952;  
Best Local Similarity 84.6%; Pred. No. 6.6e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 cggcgcgcctattattgcaggtattc 30  
||||| ||||| ||||| |||||  
Db 747 CGCGCGCCCTTTCTTGCACTTATTC 772

RESULT 15

AG098165  
LOCUS Pan troglodytes DNA, clone: PTB-100A23.F, genomic survey sequence.  
DEFINITION AG098165  
VERSION AG098165.1 GI:16718682  
KEYWORDS GSS; GSS (genome survey sequence).  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male BAC Library clone:PTB-100A23.F.

REFERENCE

1 (sites) Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE

BAC end sequences of library PTB

JOURNAL

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpanzee@gsr.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/; Tel: 81-45-503-9111, Fax: 81-45-503-9170).

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 952)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov  
Plate: LLM10420 row: p column: 03  
High quality sequence stop: 715.  
Location/Qualifiers  
1. .959  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4521458"  
/clone\_lib="NIH\_MGC\_92"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site: 1; NotI; Site: 2; SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."  
BASE COUNT 290 a 170 c 235 g 264 t  
ORIGIN

FEATURES

source

Query Match 61.3%; Score 19.6; DB 10; Length 959;  
Best Local Similarity 84.6%; Pred. No. 6.6e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 ggcgcgcctattattgcaggtattct 31  
||||| ||||| ||||| |||||  
Db 934 GCGCGCACTGTTATCGCAGATATTCT 959

Search completed: July 15, 2002, 21:49:29  
Job time: 18920 sec

